THE UNIVERSITY
of EDINBURGH

Spatial transcriptome analysis of the human developmental haematopoietic stem cell niche

Edie Isabel Flora Francis Crosse

THE UNIVERSITY *of* EDINBURGH

Doctor of Philosophy (PhD)

University of Edinburgh

2020

Abstract

During human embryonic development the first definitive haematopoietic stem cells (HSCs) arise in the aorta-gonad-mesonephros (AGM) region between Carnegie stages (CS) 13 to 17. More specifically, these HSCs are considered to derive from the endothelial cells of the dorsal aorta which undergo an endothelial-to-haematopoietic transition and bud into the lumen to form intra-aortic clusters of haematopoietic cells. This process is localised to the ventral domain and spatial polarisation of signals along the dorsal-ventral axis of the dorsal aorta has been shown to be essential for HSC emergence in the mouse AGM. This PhD project utilised laser capture microscopy coupled with RNA-Sequencing (LCM-Seq) to generate a spatial transcriptome map of the human embryonic AGM region and in doing so revealed many polarised genes and signalling pathways. Single cell and bulk sorted population RNA-Seq analysis of the dorsal and ventral portions of the aorta further resolved signalling in the AGM both elucidating a gene signature for haematopoietic progenitors and mapping niche signals to cell populations. Genes encoding the blood pressure regulators renin and endothelin-1 were amongst the most ventrally enriched factors with elements of their signalling pathways expressed across the niche on perivascular, endothelial, haematopoietic progenitor and mature cells. Renin, endothelin-1 and endothelin-2 were functionally assessed for their role in promoting HSC emergence and were shown to enhance HSC production in culture of E9.5 caudal tissues of the mouse embryo as validated by enhanced reconstitution of the haematopoietic system of irradiated mice. The “transcriptome map” dataset generated in this study can be utilised to identify signals that promote HSC emergence and can be a reference for improving haematopoietic differentiation cultures from pluripotent stem cells.

Lay Summary

Haematopoietic stem cells (HSCs), which are stem cells that produce all the cells of the adult blood and immune system, are first generated in the early human embryo. The aim of this project was to investigate the biological signals and processes that lead to embryonic HSC generation. Specifically, HSCs develop in the dorsal aorta, the central blood vessel which runs through the body of the embryo. In this project we have investigated the cells of the dorsal aorta to see what genes are switched on at the exact time and place of HSC generation. This analysis will help reveal the mechanisms and processes underlying HSC production in the embryo. By understanding these signals we can hope to mimic them in a cell culture system and artificially generate HSCs. When successful, these HSCs would be used in the clinic to treat leukaemia, lymphoma and other blood disorders.

Declaration

I declare that this thesis has been composed entirely by myself, been solely the result of my own work unless explicitly stated and has not been submitted for any other degree or professional qualification. Due references have been provided on all supporting literatures and resources.

Edie Crosse

Edinburgh, 2020

13/02/2020

Acknowledgements

I would first like to thank my supervisor Professor Alexander Medvinsky for giving me this project which turned out to be incredibly fascinating. Also, for his continued guidance and critical intellectual and creative input. I have been inspired by his passion for the subject and the depth of his knowledge and understanding.

I would also like to thank Drs Sabrina Gordon-Keylock and Stanislav Rybtsov for all their time, guidance and patience throughout this project, particularly with the haematopoietic cultures and transplantation experiments. I have had fantastic support from my all my lab members throughout my PhD and I recognise that they have all contributed towards this project either through help with an experiment or by contributions to discussion during lab meetings. I particularly thank Dr Antoniana Batsivari for showing me the ropes when I first joined and for her proactive attitude. I thank Dr Sara Tamagno for unfailingly being ready to help or discuss anything – so many things I couldn't see clearly were obvious to her. Thank you also to Anahí Binagui-Casas for her insights throughout the project and particularly for help setting up and co-ordinating the single cell experiments.

I would also like to thank my thesis committee members, Professor Steve Pollard and Professor Richard Anderson for their critique and for helping to shape the direction of my project.

I am grateful to work at the MRC Centre for Regenerative Medicine which has great facilities but particularly great facility staff. I would like to thank Dr Bertrand Vernay and Dr Matthieu Verment for their help with all things imaging related. I would also like to thank the Flow Cytometry facility staff, Drs Fiona Rossi and Claire Cryer for my extended training and their continued support. Also, thank you to Theresa O'Connor and the tissue culture facility staff for ensuring the smooth-running of my TC experiments.

I'd also like to thank my parents for giving me the best opportunities they could and all my siblings, particularly Jordan, for being there through the good, the bad and the ugly. Finally, thank you Mikey, for everything.

Acronyms

AGM	Aorta-gonad-mesonephros region
AML	Acute myeloid leukaemia
Ao	Dorsal aorta
AoD	Dorsal portion of the dorsal aorta
AoV	Ventral portion of the dorsal aorta
BM	Bone marrow
CFU-S	Colony-forming unit spleen
CHT	Caudal haematopoietic tissue (Zebrafish)
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
D-V	Dorsal-ventral
ECM	Extracellular matrix
EMP	Erythro-myeloid progenitor
ESC	Embryonic stem cell
FACS	Fluorescence Activated Cell Sorting
GVHD	Graft-versus-host-disease
HLA	Human leukocyte antigen
HSC	Haematopoietic Stem Cell
HSCT	Haematopoietic stem cell transplants
HSPC	Haematopoietic Stem Progenitor Cell
iPSC	Induced pluripotent stem cells
LCM	Laser capture microdissection
LCM-Seq	Laser capture microdissections and RNA-Sequencing
LPM	Lateral-plate mesoderm
MPP	Multipotent progenitor
Pre-HSC I	HSC precursor type 1
Pre-HSC II	HSC precursor type 2
Pro-HSC	Earliest defined HSC precursor
PSC	Pluripotent stem cells
RA	Retinoic acid
TF	Transcription factor
VA	Vitelline artery
WT	Wild-type

Contents

1. Introduction

1.1 **Overview**

The first long-term self-renewing haematopoietic stem cells (HSCs) which can generate all haematopoietic lineages first emerge during development within the embryo proper. In mammals, HSCs arise from endothelial cells in the dorsal aorta within a region of the embryo called the aorta-gonad-mesonephros (AGM) region (Medvinsky and Dzierzak, 1996; Medvinsky *et al.*, 1993). HSCs in the AGM are rare, and there are not currently markers which identify them from within the larger haematopoietic progenitor population which has hampered their characterisation and understanding of the mechanisms behind their emergence (Kumaravelu *et al.*, 2002; Medvinsky, Rybtsov and Taoudi, 2011). However, much progress has been made in the mouse model in revealing the complex signals in the AGM which underlie HSC generation. Yet there is relatively little known of the human AGM. This PhD project has investigated signalling within the human AGM niche, and in turn identified factors that support HSC generation with implications for ex vivo derivation of human HSCs for clinical translation.

This introduction provides the background for the current understanding of haematopoiesis in the embryo and the mechanisms behind HSC emergence. It first describes what an HSC is in the adult context and their clinical importance. It then describes the specifics of what is known about haematopoiesis in the embryo from studies in model organisms, including an in-depth review of signalling in the niche, before describing what is known in the human system. Finally, an overview of progress in the endeavour to artificially generate HSCs for therapeutic applications highlights the utility of developmental haematopoiesis to inform translational studies.

1.2 **What is a Haematopoietic Stem Cell?**

HSCs are the long-term self-renewing cells which, in the adult, reside in a specialised microenvironment within the bone marrow (BM) and give rise to all the mature cells of the blood and immune lineages throughout life. Studies into the effects of radiation on the body following the nuclear bombs of WWII led to the discovery that the haematopoietic system is particularly sensitive to radiation(Armstrong *et al.*, 2012). Soon after, experiments showed that transplantation of cells from haematopoietic tissues into irradiated mice facilitated haematopoietic regeneration and enhanced survival but investigators did not yet attribute this to the haematopoietic cells themselves(Jacobson *et al.*, 1951). The clonality of haematopoietic cells and thus the existence of HSCs or progenitors was first shown in 1963 by Becker et al., with the finding that colonies of radiation damaged donor-derived BM cells transplanted in the spleens of irradiated mice were made up predominantly of cells with the same chromosomal aberration (Becker, McCulloch and Till, 1963). Later experiments that tracked retrovirally marked HSCs demonstrated their capacity to give regeneration of all lineages of the

haematopoietic system (Lemischka, Raulet and Mulligan, 1986). The advent of fluorescence-activated cell sorting (FACS) technologies to sort cells based on cell surface markers has helped to refine the phenotype of HSCs in the BM and facilitate their characterisation (Spangrude, Heimfeld and Weissman, 1988).

Under physiological conditions, HSCs largely remain quiescent or slowly cycling and give rise to rapidly cycling progenitors which generate the cell turnover required for a homeostatic haematopoietic system (Foudi *et al.*, 2009; Wilson *et al.*, 2008). The classical paradigm of the haematopoietic system sits HSCs at the top of a hierarchy progressing through a multipotent progenitor (MPP) with less self-renewal capacity and through successive branching progenitors with progressively restricted potency until the mature haematopoietic cell is generated (Fig.1a). A critical point in this hierarchy is the first lineage bifurcation of progenitors with myeloid and lymphoid potential into common myeloid progenitors and common lymphoid progenitors respectively (CMPs, CLPs) (Kondo, Weissman and Akashi, 1997; Akashi *et al.*, 2000). In recent years this paradigm has been challenged. The first wave of evidence gave rise to a more plastic model of differentiation with several multipotent progenitors with lymphoid, myeloid or megakaryocytic priming but retaining multilineage potential (Adolfsson *et al.*, 2005; Mansson *et al.*, 2007; Cabezas-Wallscheid *et al.*, 2014; Sanjuan-Pla *et al.*, 2013). Furthermore, there was evidence that lymphoid progenitors still retained potential for some myeloid subtypes such as dendritic cells (DCs) and macrophages, indicating a branching point beyond that of the CLP (Doulatov *et al.*, 2010; Adolfsson *et al.*, 2005). Most recently, single cell analyses indicate that whilst the foetal haematopoietic hierarchy contains oligopotent progenitors, in the BM there is a more linear differentiation from a primed HSC/progenitor population with less clearly defined lineage branching points and only rare bipotent progenitors (Fig1.b) (Notta *et al.*, 2016; Paul *et al.*, 2015; Olsson *et al.*, 2016; Velten *et al.*, 2017).

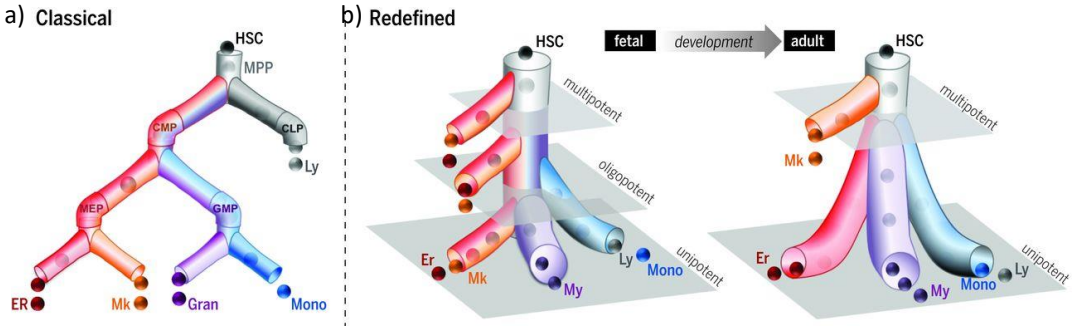


Figure 1. ***The revised haematopoietic hierarchy taken from Notta et al., 2016.*** a) The classical model shows a tree with bifurcating fate choices and stepwise lineage restriction. b) The redefined model shows some fate restriction in progenitors in the foetal system but unipotent differentiation from the stem cell compartment in the adult BM.

1.3 **The dawn of haematopoietic stem cell transplantation therapy**

The demonstrated sensitivity of haematopoietic cells to radiation in the post-war era along with the proven regenerative potential of donor haematopoietic tissues yielded the development of the first stem cell therapy – haematopoietic stem cell transplants (HSCTs). These treatments, which were initially led by E. Donnall Thomas in the 50s and for which he was awarded the 1990 Nobel Prize in Physiology, are still the most common form of stem cell treatment today. According to the 2018 report from the European Society for BM Transplantation in 2017 there were 41,100 HSCTs in Europe alone to treat, in order of frequency, lymphoid and myeloid malignancies, solid tumours and other non-malignant disorders (*European Society for Bone Marrow Transplantation - Annual Report 2018*, 2018). The patient is given radiation or chemotherapy which eradicates or depletes the cancerous haematopoietic cells but also any residual healthy clones. Then, healthy HSCs derived from the BM or umbilical cord blood (UCB) samples are transplanted into the patient where they home to the BM and reconstitute the haematopoietic system.

However, there are obstacles to be overcome. HSCTs are either autologous (patients' own cells) or allogeneic (donor derived). Allogeneic transplants need to be a close human leukocyte antigen (HLA) match which is not always possible, particularly due to the lack of ethnically diverse donor material. A poor match can lead to graft-versus-host disease (GVHD) where the immune cells of the transplant recognise the patient's cells as non-self and launch an immune attack which can also lead to liver failure (Majhail, 2017). The immunosuppressive drugs required to reduce risk of GVHD also increases the patients' susceptibility to infection and disease. Acute myeloid leukaemia (AML) survival rates for 5 years for HLA-matched and unrelated donor HSCT are ~20-60% and ~20-50% respectively, depending on how early the disease is treated (D'Souza A, 2018) .

There is a clear need for an alternative source of HSCs that provide quick reconstitution of the ablated haematopoietic system and that are well-tolerated by the patient without eliciting or executing an immune response. To this end, artificial generation of these HSCs *ex vivo* is the goal for many researchers in the field. This includes directed differentiation and reprogramming protocols to generate HSCs from embryonic stem cells (ESCs) although these cells, if poorly matched to the patient may still be immunogenic (Drukker and Benvenisty, 2004). Another strategy is to take the patients' own cells, screen them for the cancer-causing mutagens or gene-correct them, re-program them into pluripotent stem cells (PSCs) called induced PSCs (iPSCs) then direct differentiation towards HSCs. In this way, a need for a donor match would be negated. Furthermore, methods to expand HSCs in culture without losing their 'stem-ness', i.e. their multilineage potential and long-term self-renewal capacity, need to be refined. Recent reports have demonstrated great progress in the expansion of mouse HSCs *ex vivo* through fine tuning of levels of key secreted factors in culture (Wilkinson *et al.*, 2019). Furthermore, recent clinical trials have demonstrated safe engraftment of human umbilical

cord blood HSCs expanded with the cell cycle agonist UM171 (Cohen *et al.*, 2020). Further refinement of these methods would not only generate a source of HSCs for transplantation, but they would also provide a potentially limitless supply for *in vitro* drugs and therapy testing, as well as provide easily manipulatable models of disease (Daniel *et al.*, 2016). A deep understanding of the niche signals that promote HSC development *in vivo* facilitates optimisation of culture protocols that mimic this development from PSCs.

1.4 The embryonic origins of HSCs

The onset of haematopoiesis in the embryo occurs in 3 sequential waves. The first two are transitory and generate haematopoietic progenitors to meet the immediate needs of the developing embryo and subsequently the third wave generates adult-type self-renewing HSCs that power the haematopoietic system throughout life (Fig.2).

i: The primitive wave

The first wave occurs around embryonic day (E) 7.5 in the mouse and is referred to as the 'primitive wave'. This takes place in the yolk sac (YS) in mesoderm derived clusters of cells adhering to the yolk sac endoderm termed 'blood islands' (Ferkowicz and Yoder, 2005). These blood islands are the site of generation of primitive erythroid progenitors and nucleated erythrocytes which express embryonic-specific haemoglobins (Palis *et al.*, 1999). These enter the bloodstream at the onset of circulation at E8.25 and provide oxygen to the growing tissues which are now too large to receive sufficient oxygen by diffusion. During the primitive wave, macrophages and megakaryocytes are also generated (Palis *et al.*, 1999; Tober *et al.*, 2007).

ii: The second or 'pro-definitive' wave

The second wave beginning at E8.25 which also takes place in the yolk sac is defined by production of erythro-myeloid progenitors (EMPs) which generate erythroid cells expressing a unique combination of embryonic and adult-type haemoglobins along with myeloid cells (Bertrand *et al.*, 2005; Palis *et al.*, 1999; Tober *et al.*, 2007; McGrath *et al.*, 2011). Furthermore, from E9 there is production of MPPs with granulocytic, monocytic and the first lymphoid potential but without erythroid and megakaryocytic potential (Yoshimoto *et al.*, 2012; Boiers *et al.*, 2013; Palis *et al.*, 1999; Tober *et al.*, 2007). The progenitors of these restricted MPPs derive from endothelial cells with haematopoietic potential termed haematogenic (Yoshimoto *et al.*, 2012). The second wave cells are generally considered to serve the embryonic and foetal immune systems, but there is a growing body of evidence that some cells, such as EMP-derived monocytes may persist into adulthood (Ginhoux and Guilliams, 2016).

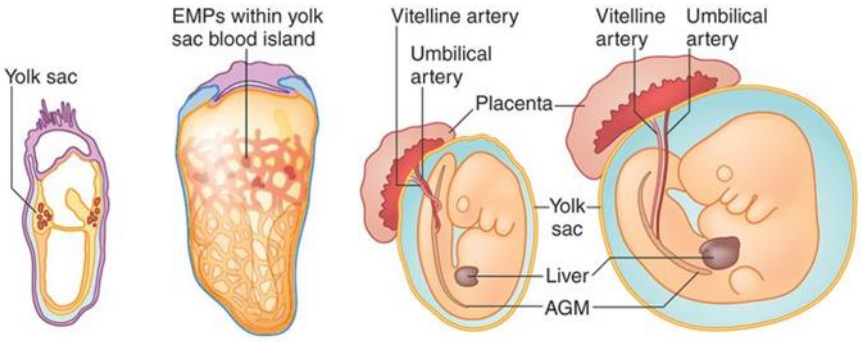
iii: The definitive wave

The third wave of haematopoiesis occurs within the embryo proper and produces adult-type HSCs termed definitive HSCs. Evidence for an intraembryonic origin of adult haematopoiesis first emerged in the 70s with seminal experiments from Dieterlen-Lièvre in the avian model (Dieterlen-Lievre, 1975). An inter-species chimera was generated comprising a quail body and chick extra-embryonic tissue which demonstrated that before hatching, the haematopoietic tissues were colonised by quail, and therefore intra-embryonic derived, cells. Later publications by Lassila et al., confirmed this observation in the adult bird (Lassila *et al.*, 1978). Prior to this, haematopoiesis was widely considered to originate solely from the yolk sac. An intraembryonic production of adult-type haematopoiesis was then confirmed in the amphibian system (Maeno, Tochinai and Katagiri, 1985; Ciau-Uitz, Walmsley and Patient, 2000; Turpen, Knudson and Hoefen, 1981). However, it was a while longer before it was demonstrated in the mammalian system. Pivotal findings came in 1993 utilising as a measure of MPP haematopoietic capacity the Colony-Forming Unit-Spleen (CFU-S) Assay whereby haematopoietic cells are injected into sub-lethally irradiated mice and macroscopic haematopoietic colonies in the spleen are measured. In this paper, CFU-S were not found in the YS prior to E9.5 despite the production of blood there since E7.5 (Medvinsky *et al.*, 1993). Furthermore, when the yolk sac begins producing CFU-S from E9.5, there is a consistently higher production in the intraembryonic region of the aorta-gonad-mesonephros (AGM) region which comprises the dorsal aorta and the urogenital ridges (UGRs). Subsequently, it was determined that HSCs arise in the AGM at E10 before the YS by using the gold-standard test of an HSC, the long-term repopulation assay (Medvinsky and Dzierzak, 1996; Muller *et al.*, 1994). This involves injection of donor cells into an irradiated recipient mouse and analysis of its capacity for reconstitution of the haematopoietic system. A definitive (adult-type) HSC is defined by its ability to provide long-term, multi-lineage reconstitution of the haematopoietic system in this assay which remains the only true read-out of an HSC.

Although, this study indicated definitive haematopoiesis is initiated in the AGM, it could not be ruled out that it is not also initiated in the YS. A novel organ culture system allowed separate culturing of AGM and YS explants from E10 and E11 embryos and showed conclusively that although both gave rise to CFU-S in culture, only the AGM has cell-autonomous generation of HSCs (Medvinsky and Dzierzak, 1996). HSCs in the AGM specifically emerge between E10.5 and E11.5 and are rare with approximately 1-3 per AGM quantified by transplanting several dilutions of low AGM cell numbers into irradiated recipient mice (Kumaravelu *et al.*, 2002; Medvinsky and Dzierzak, 1996). They then migrate to the foetal liver and undergo significant expansion (Johnson and Moore, 1975; Houssaint, 1981; Ema and Nakauchi, 2000; Kumaravelu *et al.*, 2002) before they migrate to the adult haematopoietic niches; the BM, the spleen and the thymus. In zebrafish, HSCs colonise the caudal haematopoietic tissues (CHT), which is the equivalent HSC expansion niche to the mammalian foetal liver, via the posterior

cardinal veins (Kissa *et al.*, 2008) and a similar colonisation via the blood vessels is expected in the mammalian system.

The degree of relatedness between the 3 waves of haematopoiesis is not yet fully understood. One model suggests that the 3 waves derive from the same haematogenic endothelial cell that undergoes 3 distinct spatio-temporally controlled maturations (Dzierzak and Bigas, 2018) although primitive EMPs show, at least in part, separate mesodermal origins to the later waves (Zovein *et al.*, 2010). Alternatively, the haematopoietic cells from each of the 3 waves may derive from distinct cell types. Indeed, HSCs and second wave EMPs have been shown to derive from distinct haematogenic endothelial cell populations (Chen *et al.*, 2011). It was also postulated that the cells from earlier waves are required for the specification and/or maturation of HSCs in the definitive wave and it has been shown that myeloid cells from earlier waves produce inflammatory signatures that promote HSC emergence (Espin-Palazon *et al.*, 2014; Mariani *et al.*, 2019). Fine dissection of the spatio-temporally controlled signalling events that lead to the differential development of the 3 waves will be of great importance to facilitate future *in vitro* HSC derivation cultures.



	Embryonic stage			
	E7	**E8.25**	**E9**	**E10.5**
Wave 1	Primitive erythroid progenitors macrophages megakaryocytes	Primitive erythroid progenitors macrophages megakaryocytes	Primitive macrophages	Primitive macrophages
Wave 2		EMPs Immune-restricted progenitors (persisting into adulthood)	EMPs B lymphoid cells T lymphoid cells	EMPs (yolk sac) B lymphoid cells (AGM, yolk sac) T lymphoid cells (AGM, yolk sac)
Wave 3				Hematopoietic stem cells (AGM)

Figure 2. Haematopoiesis during mouse embryonic development adapted from Yoder MC, 2014 (Yoder, 2014). Haematopoiesis in the mouse embryo occurs in three waves of increasing developmental complexity in different tissues and haematopoietic organs culminating in the production of adult-type HSCs in the AGM and extra-embryonic arteries at E10.5

1.4.1 **Other sites of haematopoiesis in the embryo**

i. *Placenta*

The placenta is also a source of HSCs in the embryo, but it remains uncertain over whether they emerge there or have migrated there from the AGM. They appear there within the same time window as HSC emergence in the AGM between E10.5 and E11.5 but soon expand to much higher numbers than in the AGM; 13 – 50 HSCs at E12.5 (Ottersbach and Dzierzak, 2005; Gekas *et al.*, 2005). The strong association of the HSCs with the endothelial cells in position and gene expression indicates derivation from a haematogenic endothelium and the placentas in Ncx1(-/-) mice which lack blood circulation generate cells with myelo-erythroid and lymphoid potential(Gekas *et al.*, 2005; Ottersbach and Dzierzak, 2005; Rhodes *et al.*, 2008). However placenta explant cultures have failed to maintain HSCs so direct evidence for autonomous production of HSCs in the placenta is lacking (Gekas *et al.*, 2005; Ottersbach and Dzierzak, 2005; Robin *et al.*, 2006).

ii. *Vitelline and umbilical arteries*

The extraembryonic vitelline and umbilical arteries which connect the dorsal aorta to the yolk sac and placenta respectively also contain haematopoietic clusters and functional HSCs as demonstrated by repopulation assays(de Bruijn *et al.*, 2000; Garcia-Porrero, Godin and Dieterlen-Lievre, 1995). Furthermore, culture of these vessels under conditions that support HSC maturation from precursors (more details in section 1.6.1) expanded the numbers of HSCs indicating that the extra-embryonic arteries are sites of HSC generation (Gordon-Keylock *et al.*, 2013).

iii. *Head*

Multilineage LT-HSCs have also been shown to emerge in the embryo head concomitantly with AGM HSC emergence at E10.5 – E11.5(Li *et al.*, 2012). However, attempts to visualise an EHT in the head were unsuccessful (Iizuka *et al.*, 2016; Li *et al.*, 2016). Therefore, HSC migration to the head from the AGM or other tissues cannot be ruled out.

1.5 **HSC development in the AGM**

Within the AGM, HSCs emerge within the dorsal aorta as demonstrated by reconstitution of irradiated mice by cells from the dorsal aorta but not the UGRs (Taoudi and Medvinsky, 2007b; de Bruijn *et al.*, 2000). Their emergence (E10.5-11.5 in mouse) is concurrent with the appearance of clusters of haematopoietic cells adhering to the ventral endothelial cells of the dorsal aorta and budding into the lumen referred to as intra-aortic haematopoietic clusters (IAHCs) (Fig.3) (Medvinsky *et al.*, 1996). IAHCs are considered to contain the emerging HSCs and haematopoietic progenitors. The close positional relationship with the endothelium and their expression of both endothelial (CD31, VE-

Cadherin) and haematopoietic (CD45) markers suggests that these cells, like the second wave haematopoietic cells, derive from the endothelial cells themselves in a process called endothelial-to-haematopoietic transition (EHT), evidence of which has been shown in several organisms (Fig.3b). In the chick, IAHCs are traced from endothelial cells by intra-aortic labelling with low-density lipoproteins before IAHCs appear (Jaffredo *et al.*, 1998). In zebrafish live imaging shows KDR+ endothelial cells change morphology becoming more rounded in shape, budding into the lumen and begin expression of transcription factors (TFs) required for haematopoiesis (Runx1, cMyb) (Bertrand *et al.*, 2010; Kissa and Herbomel, 2010). This live change in cell fate occurred without cell division but a marked morphological change and movement away from the endothelium thus demonstrating EHT. Strong evidence for mammalian EHT has been provided by ES cell cultures in which endothelial cells, derived from a mesodermal cell intermediate, were shown to lose their tight junctions, become less adherent and gain a rounded haematopoietic morphology as they concomitantly upregulated haematopoietic markers and downregulated endothelial markers (Eilken, Nishikawa and Schroeder, 2009; Lancrin *et al.*, 2009). Furthermore, in the mouse freshly cut thick transverse sections of the dorsal aorta from *Sca1*-GFP mice, which marks HSCs, showed single GFP+ cells budding from the ventral endothelium (Boisset *et al.*, 2010). The frequency of EHT events per embryo was originally estimated to be 1.7 in line with estimations for the number of HSCs in an AGM (Boisset *et al.*, 2010). However, a recent study visualising EHT in a similar way but with expression of the TF *Gata2* marking HSCs, estimated ~20 EHT events per embryo(Eich *et al.*, 2018). This higher number is more in line with IAHC numbers in the embryo and the numbers of HSC pre-cursors which are discussed in greater detail in section 1.6.

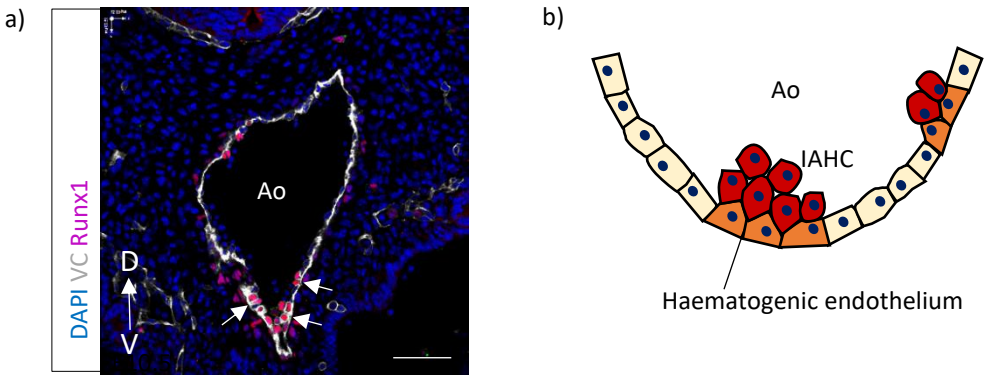


Figure 3. Intra-aortic haematopoietic clusters (IAHCs) appear in the dorsal aorta adherent to the ventral endothelium. a) IAHCs appear at E10.5 in the mouse and express endothelial (VE-Cadherin, VC) and haematopoietic markers (Runx1 and CD45). D = Dorsal, V= Ventral, Ao = Dorsal Aorta. Arrows indicate IAHCs. b) IAHCs contain Haematopoietic Stem/Progenitor Cells that emerge from haematogenic endothelium in a process called endothelial-to-haematopoietic transition.

Questions remain over how the endothelium is specified to become haematogenic. As HSC and IAHC emergence is predominantly spatially polarised to the ventral portion of the dorsal aorta, what makes the ventral endothelium haematogenic and the dorsal not? In the avian model the ventral floor and the dorsal roof of the aorta have distinct origins, deriving from the lateral and somitic mesoderm respectively, intimating that differing ancestral origins distinguish haematogenic from non-haematogenic endothelium (Pardanaud *et al.*, 1996). In the mouse, *in vivo* labelling has shown that the endothelium of the dorsal aorta and IAHCs derive from the lateral plate mesoderm (LPM) as does the endothelium of the other sites of haematopoiesis; the vitelline artery and its clusters and the placenta (Zovein *et al.*, 2010). No clear dorsal-ventral differences in mesodermal origins of the dorsal aorta suggest either a later divergence in haematogenic vs non-haematogenic fate or in situ haematopoietic competence specified by the niche. In zebrafish there is evidence of haematogenic specification in the mesoderm. The mesoderm starts upregulating endothelial (*Flk1, Fli1)* markers followed by haematopoietic markers (*Scl, Gata2, Lmo2)* even before migration below the notochord and formation of the dorsal aorta (Dooley, Davidson and Zon, 2005; Gering *et al.*, 1998). Furthermore, Notch signalling between migrating posterior lateral mesoderm cells destined to become endothelium and the somites is required for the later emergence of haematopoietic cells in the dorsal aorta suggesting that haematogenic endothelium may be specified, at least in part, en route (Kobayashi *et al.*, 2014).

1.6 The HSC hierarchy

1.6.1 HSCs in the AGM mature from pre-HSCs

There are an estimated 1-3 HSCs in the AGM as determined using a limiting dilution transplantation analysis (Kumaravelu *et al.*, 2002). Importantly, 1 HSC is sufficient to reconstitute the haematopoietic system of an irradiated recipient. Several dilutions of low cell numbers are required to quantify the numbers of HSCs, based on the ratio between repopulated and non-repopulated animals (Szilvassy *et al.*, 1990). Explant cultures of the AGM region are able to expand HSC numbers indicating their maturation from precursor cells (Medvinsky and Dzierzak, 1996). The development of a reaggregate culture system has greatly facilitated investigation into these potential precursors to HSCs or 'pre-HSCs' (Taoudi *et al.*, 2008; Rybtsov *et al.*, 2011). This involves dissociation of the AGM and either all cells are used or labelled cell populations are selected by FACS so that their fate can be tracked later. The cells are co-aggregated into balls with the AGM stromal cells or supportive OP9 stromal cells. These aggregates are then cultured on a membrane floating on pro-haematopoietic media including the cytokines SCF, IL-3, and Flt-3 which allow pre-HSCs to mature which can be then tested by transplantations using a long-term repopulation assay.

Direct transplantations into adult irradiated recipients showed that AGM HSCs are within the population of VC+CD45+ cells (Taoudi *et al.*, 2005)(Fig.4). Floating aggregate cultures of E11.5 AGMs

for 4 days dramatically increased numbers of HSCs from 1 per AGM to ~150 (Taoudi *et al.*, 2008). Analysis of the division rates of cells in the cultures showed that the majority of HSCs are dividing slowly and could not account for the large increase in HSC numbers. This demonstrated that rather than high proliferation of the single initial HSC there is recruitment of precursor cells into HSCs. These pre-HSCs, termed pre-HSC II, were also within the VC+CD45+ compartment. There has since been discovery of even earlier precursors which sequentially develop into HSCs called, in order of development, pro-HSC and pre-HSC I (Rybtsov *et al.*, 2011; Rybtsov *et al.*, 2014). Pre-HSC I were discovered through the observation that although E10.5 AGMs do not reconstitute irradiated mice in the repopulation assay, indicating a lack of HSCs, 4 day *ex vivo* aggregate culture generated HSCs from the VC+CD45- fraction (Rybtsov *et al.*, 2011). These cells were subsequently shown to mature into HSCs via a pre-HSC II intermediate and express at low levels the haematopoietic marker CD41, distinguishing them from the VC+CD45- endothelial cells. Pro-HSCs emerge at E9.5 and required a longer culture period of 7 days in order to derive HSCs from them consistent with their development through pre-HSC I and pre-HSC II intermediates (Rybtsov *et al.*, 2014). These cells are VC+CD45-CD41lo and are distinguishable from pre-HSC I by the lack of CD43 expression.

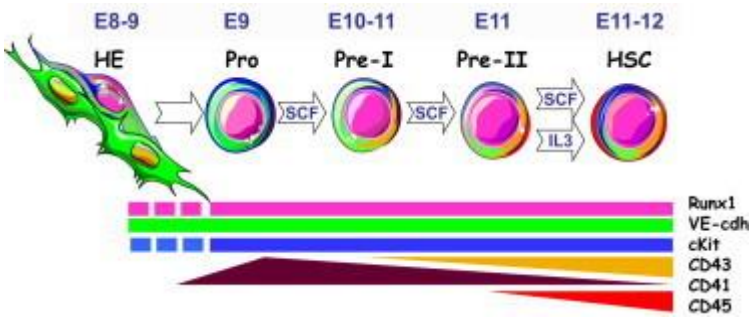


Figure 4: Murine HSCs develop from haematogenic endothelium (HE) through stepwise production of precursors. Taken from Rybtsov et al, 2014. A schematic of HSC development shows endothelial-to-haematopoietic transition through precursor intermediates; pro-HSCs at E9 which upregulate CD41 but lack CD43/CD45 expression, pre-HSC I at E10-11 which have low CD41 expression but begin to express CD43 and finally, pre-HSC II which upregulate expression of CD45 are immunophenotypically similar to HSCs which appear at E11. Runx1, VE-Cadherin, and cKit are expressed throughout.

1.6.2 **Quantification and dynamics of pre-HSCs**

Limiting dilution analyses quantified the numbers of pre-HSCs through HSC development identifying only 1 or 2 pro-HSCs at E9.5, 20-30 pre-HSCs at E10.5 and 65 at E11.5, dropping down to 10 at E12.5 (Rybtsov *et al.*, 2016). Notably, the extra-embryonic vitelline and umbilical arteries also contain pre-HSCs capable of maturing into adult-type HSCs (Gordon-Keylock *et al.*, 2013). The foetal liver does not contain any pre-HSCs further demonstrating that HSC maturation is localised to the AGM and extra-

embryonic vessels prior to migration to the foetal liver (Rybtsov *et al.*, 2016). The expansion of pre-HSCs in the AGM region is more in line with the numbers of definitive HSCs (50-60) found in the foetal liver than the 1-3 AGM HSCs which would have to undergo intensive proliferation within too short a time (Kumaravelu *et al.*, 2002; Ema and Nakauchi, 2000). These proliferation dynamics were visualised with Fucci dual reporter mice which differentially label cells in G_0/G_1 and $S/G_2/M$ phases (Batsivari *et al.*, 2017). Pro-HSCs at E9.5 arise from actively cycling endothelial cells but are slowly cycling themselves. At E10.5 and 11.5 the pre-HSC I and II populations are both actively cycling which corresponds to the expansion seen at this stage. Finally, HSCs are slowly cycling, corresponding to their low number in the AGM. Single cell analyses of the E11 pre-HSC populations resolved their dynamics further, demonstrating populations of actively proliferating and relatively quiescent pre-HSCs which were CD168+ and CD168- respectively. Notably the proliferating cells gave higher chimerism reconstitution *in vivo* indicating that proliferating pre-HSCs are predisposed to become HSCs (Zhou *et al.*, 2016).

1.7 Dynamic gene expression across the EHT

1.7.1 Master transcription factors specifying HSC generation

There are several transcription factors (TFs) involved in specifying HSC generation from the haematogenic endothelial cells and various combinations of these have been used to reprogram cells into HSCs (discussed in greater detail in section 1.10.2). The TFs described in this section are considered master regulators due to their critical roles upstream of many other haematopoietic TFs and their loss of function results in serious haematopoietic defects.

The earliest specification towards adult-type HSC generation occurs in the lateral plate mesoderm (LPM). These TFs include FLI1, ETV2, GATA2 and SCL/TAL1. In the Xenopus model, these TFs are responsible for specification of the LPM to become hemangioblasts; cells which will later migrate to the midline to form the dorsal aorta and give rise to haematogenic endothelium and HSCs (Ciau-Uitz *et al.*, 2013). Fli1 and Etv2 induce expression of Gata2 which in turn activates the expression of the vascular endothelial growth factor (VEGF) receptor Flk1. This allows cells to become responsive to VEGFA signalling, a pathway required for activation of SCL expression and, in turn, specification of the hemangioblast (Ciau-Uitz *et al.*, 2013; Ciau-Uitz *et al.*, 2010). In the mouse, SCL patterns the mesoderm to become blood-forming extraembryonic and LPM over paraxial mesoderm (Ismailoglu *et al.*, 2008).

During migration of the hemangioblast cells to the midline to form the dorsal aorta SCL and Gata2 expression is reported to be lost (Ciau-Uitz, Walmsley and Patient, 2000). However, they both have additional critical roles in the generation of HSCs. In zebrafish, one isoform of SCL, *scl-β* was required for formation of the haematogenic endothelium whereas the second isoform, *scl-α* was required for the maintenance of HSCs after EHT (Zhen *et al.*, 2013). *GATA2*$^{-/-}$ human embryonic stem cell (ESC)

culture directed to HSC fate was used to model its effects on EHT and found that *GATA2* was not required for the haematogenic endothelium specification but is a critical regulator of EHT (Kang *et al.*, 2018). *In vivo* imaging of thick transverse sections of the dorsal aorta provided live imaging of *Gata2* cells undergoing EHT (Eich *et al.*, 2018; Nottingham *et al.*, 2007). Gata2 is also required for the continued survival of HSCs following EHT (de Pater *et al.*, 2013).

Gata2 and SCL act upstream of the TF Runx1 (Nottingham *et al.*, 2007). Runx1 is a well-studied TF with an essential role in the generation of HSCs. It is expressed in the mesoderm of the murine YS blood islands as well as the haematogenic endothelium and IAHCs of the dorsal aorta and extra-embryonic vessels (North *et al.*, 1999) which corresponds with its involvement in specification of both haematopoietic progenitors and HSCs from endothelial cells. *Runx1* knock out resulted in an absence of IAHCs in the dorsal aorta suggesting that is involved in EHT (Yokomizo *et al.*, 2019; Yokomizo *et al.*, 2001; North *et al.*, 1999). This has been demonstrated with live imaging in zebrafish where a morpholino knock down (KD) of *Runx1* inhibited EHT and any cells that did attempt to emerge from the aortic endothelium were abortive and disintegrated soon after (Kissa and Herbomel, 2010). A conditional knock out (cKO) of *Runx1* in the dorsal aorta endothelium of mice prevented HSC emergence suggesting a similar critical role in EHT regulation (Chen *et al.*, 2009). However, unlike Gata2, Runx1 was not necessary for cell survival following specification and HSC formation. Runx1, has been shown to regulate not only key haematopoietic genes but also those involved in cell adhesion and migration, a key part of the EHT process(Lie *et al.*, 2014).

The above factors form part of a network of highly interconnected TFs. Fli1, despite being expressed later than SCL during directed ES cell differentiation towards haematopoietic cells, has been shown to induce expression of SCL as well as Runx1, Gata2 and other HSC specification TFs Lmo2 and Etv2 putting it at the top of the HSC specification hierarchy (Liu *et al.*, 2008; Lichtinger *et al.*, 2012). *Runx1* expression in HSCs is controlled by an enhancer region that recruits Fli-1, SCL and Gata2 (Nottingham *et al.*, 2007). Interesting dynamics in binding patterns of the TFs have been shown in an ES culture model of EHT (Lichtinger *et al.*, 2012). In response to *RUNX1* expression, SCL and Fli1 shift their enhancer binding patterns in the haematogenic endothelium to the haematopoietic precursor cells where they bind the motifs of genes involved in HSC specification including GATA2 and RUNX1 sites. RUNX1 then binds the original SCL and FLI1 binding sites, which it was proposed, have been primed for rapid activation by RUNX1. This highlights the finely controlled and inter-woven regulation by these master TFs in the specification of HSCs from their mesodermal origins.

1.7.2 **Acquisition and loss of arterial cell fate**

HSC development in the embryonic blood vessels occurs exclusively in the arteries and not in the veins (Taoudi and Medvinsky, 2007a; de Bruijn *et al.*, 2000). The relationship between arterial fate and HSC emergence has been shown to be closely coupled as several upstream regulators of arterial over

venous identity including Notch and Hedgehog signalling are required for HSC specification(Gering and Patient, 2005; Burns *et al.*, 2005). Similarly, Sox17 expression is required for both arterial specification and HSC formation (Clarke *et al.*, 2013; Corada *et al.*, 2013). However, although arterial specification is required for EHT, the arterial fate program must then be repressed in order to allow transition towards the haematopoietic fate, visualised by a Sox17 downregulation in the IAHCs(Lizama *et al.*, 2015). Sox17 was found to directly bind *Runx1* and *Gata2* upstream 5' untranslated regions, inhibiting their expression. Conversely *Sox17* deletion induced at E9.5 at the onset of EHT increased *Runx1* and *Gata2* expression and reduced other arterial markers such as *Dll4*. Downregulation of Notch has also been shown to be important for EHT (Souilhol *et al.*, 2016b). Furthermore, it was shown recently that the regulated mRNA decay of arterial genes *Notch* and *rhoca* allows progression to haematopoietic fate (Zhang *et al.*, 2017).

1.7.3 **Single cell analyses reveal EHT dynamics**

The explosion in single cell RNA-Seq analyses in recent years has facilitated the investigation into the molecular underpinnings of EHT. A single cell analysis of E11 AGM EHT populations (Endothelial cells, pre HSC I/II, E12 and E14 foetal liver HSCs) revealed enrichment of the mTOR pathway in pre-HSC I, which regulates cell metabolism, growth, proliferation and survival (Zhou *et al.*, 2016; Laplante and Sabatini, 2009). It was subsequently shown that endothelial disruption of the core mTORC2 component, Rictor, prevented HSC emergence from AGM explant cultures. It was also found that pre-HSC II expressed the highest amount of long non-coding RNAs (lncRNAs). The same group recently performed single cell analyses of the EHT populations but focused specifically on the differential expression of lncRNAs (Zhou *et al.*, 2019). They found highly distinctive lncRNA expression profiles resolved to each population and through a functional screening determined H19 lncRNA as a key mediator of the transition from endothelial cells to pre-HSC I. The heightened methylation levels at the promoters of key HSC development genes in *H19* mutants led the authors to suggest the lncRNA helps regulate EHT via demethylation of haematopoietic program promoters. These studies highlight the power of single cell analyses to inform on the multi-layered mechanisms of HSC.

Another single cell dataset was acquired from multiple sorted cell populations across the EHT at E10 and E11 (Baron *et al.*, 2018). Hierarchical clustering of TF expression across the dataset highlighted TF dynamics across the EHT pseudotime including endothelial *Sox17* and *Hey2* expression, haematogenic endothelial *Elk3*, *Jun* and *Mecom* expression, EHT or pre-HSC TFs *Gfi1, Runx1, Ikzf2* and lastly haematopoieis and differentiation TFs including *Myb, Zpfm1* and *Lyl1.* Another single cell study of the EHT populations in the mouse AGM elucidated CD44 as a new marker of EHT with low expression on pro-HSCs and pre-HSC I and higher expression on pre-HSC II (Oatley *et al.*, 2020). A recent study profiled single cells from the early human dorsal aorta which also indicated CD44 as an enrichment marker for the human haematogenic endothelium (Zeng *et al.*, 2019).

1.8 Signalling in the aorta-gonad-mesonephros region

During embryonic development, HSCs emerge and mature under the influence of differentiation cues from an environment that is in a constant state of growth and morphological change. Rather than receiving niche signals from mature neighbouring cells as is the case in many adult stem cell niches, the embryo stem cell is likely to be in the vicinity of many other stem cells with alternative lineage trajectories which also require a unique set of cues. These cells must act synergistically with one another to form the dynamic structures of the developing embryo. To achieve this high level of co-ordination each cell must have an exact positional and temporal identity. Often, in the embryo this is achieved with overlapping signalling gradients across tissues and the intersection of broad and local signals. The result is that the developmental stem cell niche is often complex, multi-layered and overlapping with other niches. This sub-section explores what is known about signalling in the AGM and how these pathways intersect

1.8.1 Polarisation of the niche.

As previously mentioned, HSC emergence in the AGM is polarised to the ventral domain of the dorsal aorta both in mice and in humans (Ivanovs *et al.*, 2014; Taoudi and Medvinsky, 2007a). Efforts to determine how these spatial cues are provided revealed the important role of polarised signalling from the tissues surrounding the aorta (Peeters *et al.*, 2009). Transplantation of the AGM with the ventral tissues, including the gut, generated higher CFU-S numbers than the AGM alone. Conversely, co-transplantation with the dorsal tissues significantly reduced CFU-S numbers. Correspondingly, only early E10 AGM explants cultured with the ventral tissues gave repopulation following transplantation, compared to AGM alone and AGM cultured with the neural tube (Peeters *et al.*, 2009).

Signalling from the niche is more complicated than supportive ventral signals and inhibitory dorsal signals however. An elegant study from Souilhol et al., demonstrated how the interplay of signals from the dorsal, ventral and urogenital ridges fosters an enhancing environment for HSC generation (Souilhol *et al.*, 2016a). The dorsal and ventral portions of the dorsal aortae from GFP and wild-type (WT) mice were sub-dissected and cultured together as aggregates in different combinations; ventral/ventral, dorsal/dorsal or ventral/dorsal and transplanted intro irradiated mice. The GFP allowed relative contributions of each subdomain to be tracked. Notably, although the majority of HSC contribution came from the ventral domain, at E10.5 this effect was greatly enhanced by the presence of the dorsal domains compared to culture of ventral alone. However, this effect was no longer evident by E11.5 (Fig.5). Furthermore, culturing with UGR gave enhanced HSC production than without.

The same study defined some polarised signalling molecules across the AGM and proposed a model where the intersection of their gradients creates an optimal zone for HSC emergence in the ventral aorta (Souilhol *et al.*, 2016a) (Fig.5). This includes a gradient of Stem Cell Factor (SCF) from the ventral

domain which, in culture, was able to stimulate increased HSC production from the dorsal domain. SCF has also previously been shown to support maturation of HSCs from E9 pro-HSCs (Rybtsov *et al.*, 2014). Sonic hedgehog (Shh) was dorsally polarised, as can be expected from its production in the notochord located dorsally above the aorta and could enhance HSC generation from the ventral domain. BMP signalling and downstream effector P-Smads,1,5 and 8 were ventrally enriched in the sub-aortic-mesenchyme but downregulated in the clusters and in the progression from pre-HSC I to pre-HSC II. Shh induced Noggin, a BMP inhibitor providing a potential link between these pathways in regulation of HSC emergence. Intriguingly, two types of HSCs are found in the AGM – BMP-activated HSCs and non-BMP activated/Hedgehog regulated HSCs – highlighting the complexity of signal control in the microenrvironment (Crisan *et al.*, 2016). Such gradients of signalling are a classic example of how cells in the embryo are given their identity in space and time.

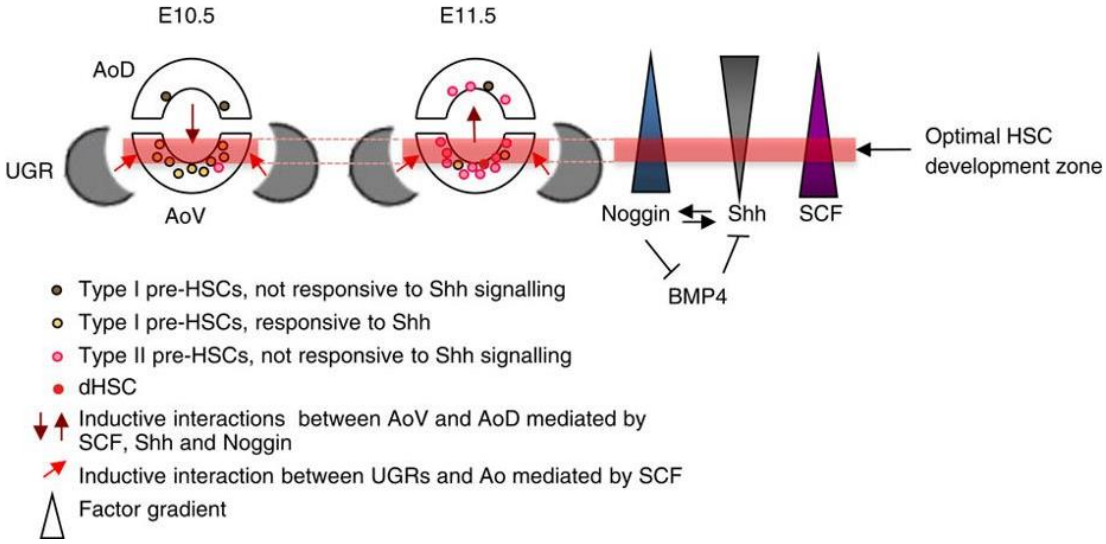


Figure 5 Polarised signalling in the mouse AGM region, taken from (Souilhol *et al.*, 2016a). HSC emergence is localised to the ventral domain of the dorsal aorta but signalling from the dorsal domains and urogenital ridges (UGRs) enhances HSC emergence. This is a proposed model of how polarised signalling pathways in the AGM may intersect to create an optimal development zone for HSCs in the ventro-lateral region of the dorsal aorta.

A comprehensive investigation into polarised signalling in the mouse AGM niche at E9.5-E10.5 has recently been carried out through RNA-Seq analysis of the sub-dissected ventral and dorsal domains of the dorsal aorta (AoV and AoD respectively) and UGRs (McGarvey *et al.*, 2017). The spatially resolved transcriptome elucidated genes and signalling pathways polarised to each of these domains. There were several pathways enriched in the E10.5 AoV as compared with E9.5 AoV (to indicate signals required for pro-HSC maturation) and the E10.5 AoD. These included pathways with already implicated signalling roles in the niche including pro-inflammatory pathways, SCF/Kit, VEGF and TGF-β. In line

with the above mentioned ventral enrichment of BMP effectors but down-regulation in the IAHCs (Souilhol *et al.*, 2016a) BMP signalling was ventrally enriched in this dataset including both BMP ligands and inhibitors. Critically, one of these inhibitors BMPER, enhanced HSC generation in culture as demonstrated with a repopulation assay. This is a good example of how spatially resolved data can be used to inform on important niche signals supporting HSC generation.

1.8.2 **Signalling in the HSC developmental niche**

The studies described in the above section highlighted some of the signalling molecules in the HSC niche which interact to enhance HSC production, including SCF, Shh, Noggin and BMP4 (Rybtsov *et al.*, 2014; McGarvey *et al.*, 2017; Souilhol *et al.*, 2016a). Additional key signalling pathways have also been described and interesting, tightly regulated mechanisms of action continue to be revealed (Fig.6).

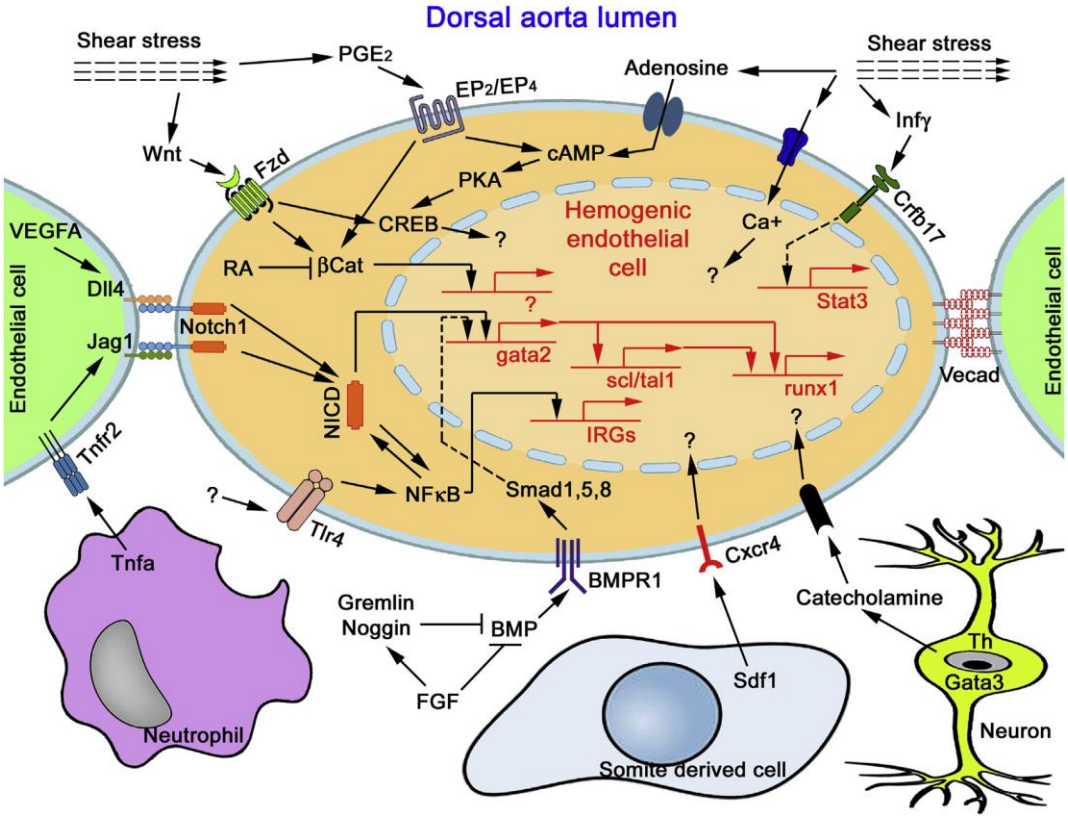


Figure 6. Signalling in the HSC developmental niche taken from (Ciau-Uitz A, 2016). Schematic of the complex, highly interactive signalling within the AGM that act synergistically to activate key transcription factors and promote HSC emergence. This includes Notch, Wnt and inflammatory signalling pathways, signalling from catecholamines and hormones as well as signals initiated by blood flow shear stress forces through the dorsal aorta lumen.

i. *Notch*

As discussed previously the Notch family members are involved in specification of arterial over venous fate (Lawson *et al.*, 2001) but must be down-regulated for HSCs to emerge (Zhang *et al.*, 2017; Souilhol *et al.*, 2016b). Notch1 expression is downregulated from pre-HSC I to II and further in HSCs. Correspondingly, co-culture with OP9s with active Notch signalling impaired maturation of pre-HSC I to II (Souilhol *et al.*, 2016b). Taken together this highlights how haematogenic endothelium specification is de-coupled from pre-HSC to HSC maturation through precisely controlled regulation of signalling molecules. Notch signalling is a central pathway within the HSC specification process at multiple stages and, as demonstrated below, is highly connected with other niche signalling pathways.

ii. *Wnt signalling*

Wnt signalling is another pathway with documented roles at several time points along the HSC production pathway demonstrating both temporal and Wnt-family member specific regulation. In zebrafish, Wnt16 was shown to be required for the production of HSCs in the dorsal aorta, but intriguingly, the mechanism of action took place before the dorsal aorta had even been formed (Clements *et al.*, 2011). Wnt16 acts through the non-canonical signalling pathway triggering Notch signalling in the somites which acts on vascular precursors as they migrate across the somites bestowing them with the ability to become HSCs in the dorsal aorta (Kobayashi *et al.*, 2014). This reinforces the idea that HSCs are at least in part specified before EHT and highlights another spatio-temporally defined role of Notch signalling. In mouse, Wnt signalling also has a role in specifying HSCs from the endothelium itself, acting through its canonical pathway effector β-catenin at E10.5 but with little or no effect by E11.5 (Ruiz-Herguido *et al.*, 2012). Although in that study Wnt was acting specifically on the endothelial cells and not the HSC progenitors, a zebrafish study showed that independently of HSPC specification Wnt9a amplifies HSPCs in the dorsal aorta prior to migration to the CHT (Grainger *et al.*, 2016).

iii. *Hedgehog signalling*

Hedgehog (Hh) signalling is also required at multiple stages in the early development of HSCs from the mesoderm. It is required for the migration of mesodermal endothelial progentiors to the midline to form the dorsal aorta (Gering and Patient, 2005). It is then also required for specification of the endothelial cells to become arterial which, with VEGF, act upstream of Notch (Lawson *et al.*, 2001). Although Hh signalling from the dorsally located notochord counteracts a ventrally derived Bmp4 signal to maintain dorsal-ventral polarity in the dorsal aorta (Wilkinson *et al.*, 2009), there is evidence to suggest that gut-derived sonic hedgehog (Shh) may also promote HSC emergence (Peeters *et al.*, 2009).

iv. *Inflammatory Signalling*

Pro-inflammatory signalling pathways also have a key role in supporting HSC development. Several reports indicate the involvement of a range of cytokines - small peptides that are usually secreted in response to infection and tissue damage to activate and mobilise immune cells (Dinarello, 2000). Primitive neutrophils provide a source of Tumor necrosis factor α (TNFα) which binds to the cell receptor TNFR2 on endothelial cells and stimulates EHT in the zebrafish dorsal aorta (Espin-Palazon *et al.*, 2014). This process was upstream of Notch, acting via the Notch ligand Jag1.

Type I and II Interferons, which are another class of potent inflammatory cytokines, also play a role. IFN-α and IFN-γ in mouse and IFN-γ and IFN-φ in zebrafish have been shown to support HSC emergence (Sawamiphak, Kontarakis and Stainier, 2014; Li *et al.*, 2014). In zebrafish this regulatory role was shown to control EHT, acting this time downstream of Notch (Espin-Palazon *et al.*, 2014). This seems to suggest that Notch has a role mediating the influence of inflammation in the HSC developmental niche. The studies of the role of TNFα and Interferons indicated that they are produced by myeloid cells deriving from the primitive haematopoietic wave, more specifically neutrophils in the case of TNFα, showing that primitive haematopoietic cells are also an integral part of the definitive HSC niche (Sawamiphak, Kontarakis and Stainier, 2014; Li *et al.*, 2014). Furthermore, macrophages interact with emerging intra-aortic clusters and may provide important pro-inflammatory signals (Mariani *et al.*, 2019).

v. *Retinoic Acid*

In mice, retinoic acid (RA) signalling has been shown to have a critical role in supporting HSC generation as only HSCs with active aldehyde dehydrogenase activity, which synthesises RA, gave repopulation in transplanted mice and E11.5 AGMs with inhibited RA signalling in the haematogenic endothelium gave no repopulation(Chanda *et al.*, 2013). This signalling effect was realised through inhibition of Wnt, highlighting once again the interconnected nature of signalling pathways in the HSC niche. Furthermore, a study which added RA receptor agonist to differentiating hESC cells found an increase in expression of arterial markers (DLL4, HEY2, SOX17) in hESC-HSPCs (CD45+CD34+CD90+) (Dou *et al.*, 2016). At low levels, RA enhanced HSPC production from human pluripotent stem cells whereas high RA blocked HSPC production suggesting RA signalling levels must be tightly controlled (Ronn *et al.*, 2015).

vi. *Shear stress and NO signalling*

In contrast to a flat cell culture an embryo has a complex tissue architecture with multiple forces and tensions acting on cells. These varied forces have a well-documented influence on cell fate(Vining and Mooney, 2017). The cells of the dorsal aorta are subjected to an array of forces including the shear stress of blood flow and the tensile, contractile forces generated by the perivascular cells of the vessel

wall and interactions with the extracellular matrix. There is a growing body of evidence to suggest that these forces inform and influence HSC emergence, with a focus on the haemodynamic forces generated by the flow of blood through the vessels. In mice and zebrafish the shear-stress forces of blood flow through the dorsal aorta have been shown to promote HSC emergence via nitric oxide signalling (Wang *et al.*, 2011; North *et al.*, 2009; Adamo *et al.*, 2009). Intriguingly, Notch has a described role as a mechanosensor to arterial over venous blood flow which, in turn, maintains arterial identity (Weijts *et al.*, 2018). It has been proposed that the haematogenic endothelium senses shear stress through cilia structures as these transduce Notch signalling to endothelial cells facilitating HSPC specification(Liu *et al.*, 2019). Furthermore, shear-stress has been shown to promote HSC emergence through not only Notch signalling but other pathways with known documented pro-haematopoietic roles including Wnt and prostaglandin signalling, placing mechanosensing high in the hierarchy of signalling in the HSC niche (Diaz *et al.*, 2015).

vii. *Signalling neuropeptides and hormones*

The sympathoadrenal progenitors found close to the dorsal aorta in ventro-lateral positions support HSC development through the release of neurotransmitter catecholamines (Fitch et al., 2012). These cells of the developing sympathetic nervous system express *Gata3* which acts upstream of tyrosine hydroxylase, an enzyme required for the biosynthesis of catecholamines. It was proposed that the catecholamines may act directly on aortic HSCs, as they express the β2-adrenergic receptor, supporting their emergence from the haematogenic endothelium. In zebrafish another neurotransmitter, serotonin, has been shown to promote HSPC formation via a cascade of secreted hormones beginning in the brain called the hypothalamic, pituitary, adrenal (interrenal) axis signalling to gluococorticoid receptors on the haematogenic endothelium and HSPCs (Kwan *et al.*, 2016). A zebrafish screen also identified the hormone-like prostaglandin E2 (PGE2) as a potential promoter of HSC formation by a demonstrated increase of *runx1+/cmyb+* AGM following incubation with a long-acting derivative of PGE2 and a decrease following incubation with an inhibitor (North *et al.*, 2007). A mechanism of action whereby PGE2 exerts its influence on HSPC emergence via Wnt signalling was shown in a later publication (Goessling *et al.*, 2009). PGE2 expanded the number of HSCs derived from mouse BM as shown by a competitive repopulation assay, highlighting how lessons from development may be of clinical use for expanding HSCs for treatment (North *et al.*, 2007).

viii. *Neural Crest*

There is an emerging role for the neural crest cells in the support of HSC generation in the dorsal aorta. The neural crest are multipotent cells originating in the neural tube which give rise to a diverse array of cell types including sympathetic and parasympathetic neurons, medulla cells of the adrenal gland, pigment cells and head skeletal and connective tissues (Gilbert, 2000). In zebrafish it was discovered that the neural crest cells migrate ventromedially and associate with the dorsal aorta just before the

onset of haematopoiesis in the dorsal aorta (Damm and Clements, 2017). This association was mediated through platelet-derived growth factor (pdgf) signalling and disruption of the association using morpholinos against neural crest migration genes impaired HSPC emergence from the haematogenic endothelium.

This section illustrates the diverse signalling in the AGM niche that comes together to create the correct inductive environment for HSCs to emerge including a range of cell types and signalling centres. It is clearly not a trivial task to be able to adequately recreate these conditions to generate HSCs *ex vivo* but a full understanding of the niche, particularly of those factors at the top of the signalling hierarchy, will facilitate this. The application of our current understanding of the developmental niche has already informed on and improved culture conditions for derivation of HSCs from other cell types, which is discussed in greater detail in sub-section **1.10**.

1.9 HSC development in the human embryo

Due to the relative scarcity of human embryos and legal restrictions in some countries, the development of HSCs in the human system is comparatively less well studied than in mouse and zebrafish. However, important work has shown that many of the processes behind HSC development in the model organisms are conserved in humans. There are however some critical differences which make the specific study of human developmental haematopoiesis of great fundamental and clinical importance. This sub-section details the current state of our understanding in the field.

1.9.1 Human HSC emergence occurs in the AGM and is polarised to the ventral domain of the dorsal aorta.

As in the mouse, the first wave of haematopoiesis is initiated in the YS blood islands at Carnegie Stage (CS) 7-8 (16 – 18.5 days post-conception (dpc)) producing primitive erythrocytes and a few macrophages and megakaryocytes (Tavian *et al.*, 1996; Tavian, Hallais and Peault, 1999; Bloom and Bartelmez, 1940; Fukuda, 1973; Luckett, 1978) (Fig.7). The blood islands are present until CS11 (24 dpc) at which stage primitive erythroid cells expressing glycophorin-A are found in the embryo proper indicative of the onset of blood circulation (Tavian, Hallais and Peault, 1999). IAHCs in the dorsal aorta appear later, first as a few scattered CD34+ cells at CS12 (27 dpc) growing in size to peak at CS14/15 (30 – 33 dpc) before declining in size again until CS16/17 (40 dpc) (Tavian, Hallais and Peault, 1999). The IAHCs were exclusively located in the ventral domain, in contrast to the mouse which has a few dorsally located IAHCs (Taoudi and Medvinsky, 2007b; Taylor, Taoudi and Medvinsky, 2010; Yokomizo and Dzierzak, 2010). Large clusters were also seen in the vitelline artery (VA) and particularly at the bifurcation of the dorsal aorta with the VA (Tavian, Hallais and Peault, 1999). These IAHCs express markers of haematopoiesis including CD45+, CD43+, the TFs c-myb, Gata2, Gata3, SCL/TAL1 as well as endothelial markers CD34 and VC (Tavian *et al.*, 1996; Tavian, Hallais and Peault, 1999; Watt *et al.*,

2000; Labastie *et al.*, 1998). The close relationship and association with the dorsal aorta suggest that HSCs arise through EHT. However, imaging of an IAHC sometimes shows a disrupted endothelium at the base of the IAHC and an association with the sub-aortic mesenchyme. This could be indicative of a mesenchymal contribution to the IAHCs but does not exclude the possibility that the cells pass through an endothelial intermediate (Tavian, Hallais and Peault, 1999). Endothelial cells sorted from the dorsal aorta gave rise to haematopoietic colonies from CS12-13 (28dpc) indicating the presence of haematogenic endothelium(Oberlin *et al.*, 2002). Furthermore, directed differentiation cultures towards an HSC-like fate indicates progression through an endothelial-like intermediate suggestive of EHT (Pereira *et al.*, 2013; Ayllon *et al.*, 2015; Ditadi *et al.*, 2015; Ronn *et al.*, 2015; Slukvin, 2013).

The AGM as the site of emergence of HSCs was confirmed by transplantation into irradiated NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/Sz (NSG) mice (Ivanovs *et al.*, 2011). The embryonic haematopoietic tissues from CS12 (26 days post-conception (dpc)) to CS17 (41 dpc), including the AGM, YS, liver, umbilical cord and placenta, were assayed in this way for the presence of HSCs. Between Carnegie Stages (CS) 14 – 17, HSCs were present in the AGM providing long-term multilineage repopulation. A small number of mice repopulated from the YS in the later stages at CS16 and 17. The foetal liver first gave repopulation at CS17 but this was uni-lineage T-cell repopulation indicating that true adult-type HSCs have not yet migrated here at this stage. Furthermore, there were no HSCs present in the umbilical cord or placenta until later stages as demonstrated by long-term repopulation from a 15 week old placenta. Taken together, HSCs emerge in the AGM around 5 days earlier than any other haematopoietic tissue consistent with it being the site of HSC initiation. Further sub-dissections identified that HSC generation occurs within the dorsal aorta, between the rostral bifurcation of the aorta and the umbilical arteries and localised to the ventral domain (Ivanovs *et al.*, 2011; Ivanovs *et al.*, 2014; Easterbrook *et al.*, 2019). In the mouse, HSC emergence has also been localised to the central third of the AGM (Mascarenhas *et al.*, 2009). Polarised specification of HSCs in the dorsal aorta along the rostral-caudal and dorsal-ventral axes is therefore conserved across species.

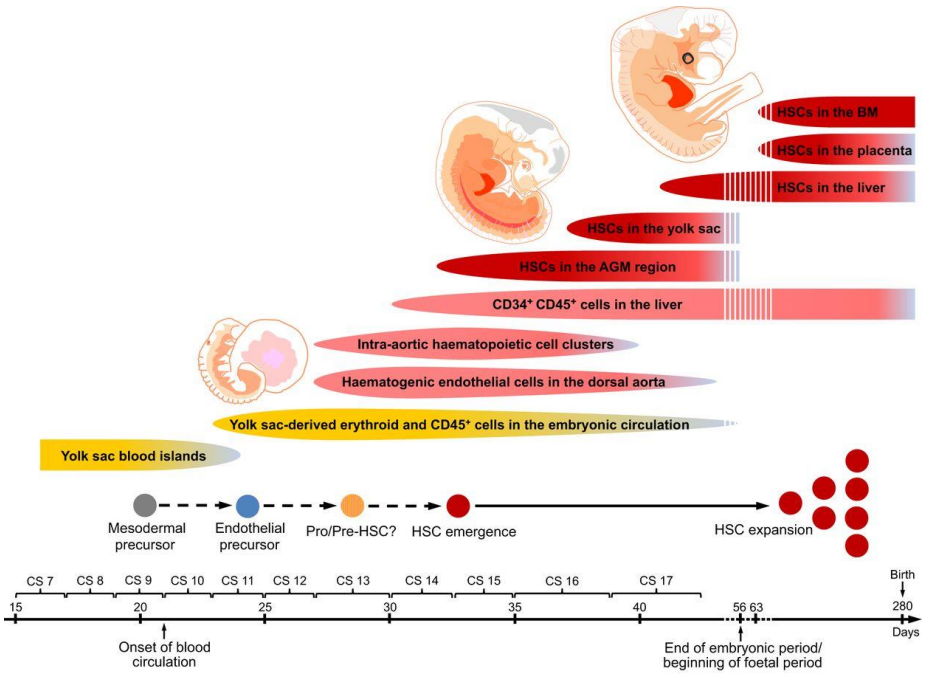


Figure 7. Timeline of haematopoiesis in the human embryo adapted from (Ivanovs *et al.*, 2017). Human embryonic haematopoiesis initiates in the yolk sac blood islands around Carnegie stage (CS) 7 – 8 (~18 days post conception, dpc). Yolk sac erythroid and CD45+ cells enter the circulation around CS11 (24dpc). Between CS12 and CS13 (~27-29dpc) haematogenic endothelial cells are present in the dorsal aorta concurrent with the appearance of intra-aortic haematopoietic clusters (IAHCs). This is just earlier than the emergence of HSCs in the AGM which begins at CS14 (~30dpc). HSCs emerge in the AGM earlier than in the yolk sac, liver and placenta.

1.9.2 The phenotype and properties of HSCs in the human AGM

There is no set of markers that exclusively identify HSCs in the AGM separately from lower potency progenitors. Efforts to refine and whittle down the population in which HSCs reside are therefore of great importance for investigation into their defining characteristics. The most comprehensive immunophenotype analysis of HSCs in the AGM defines them as $CD34^{+}VC^{+}CD45^{+}KIT^{+}THY1^{+}endoglin^{+}RUNX1^{+}CD38^{-/lo}CD45RA^{-}CD43+$ (Ivanovs *et al.*, 2014). A human AGM has approximately only one HSC although these cells have a large regenerative capacity with one AGM HSC giving rise to at least 300 daughter HSCs as demonstrated by serial transplantations with limiting dilutions (Ivanovs *et al.*, 2011). UCB samples do not have such high potency (Ivanovs et al., submitted). This regenerative potential makes generation of AGM-like HSCs an attractive target for clinical applications.

Notably, the phenotype of human AGM HSCs is not the same as in the mouse. As previously discussed, HSCs in the mouse AGM emerge from the haematogenic endothelium via precursor intermediates which sequentially upregulate expression of CD41, CD43 and CD45. Although, human AGM HSCs do express CD45 and CD43, CD41 does not capture the HSC compartment (Ivanovs et al., unpublished).

The differences between mouse and human are exemplified through the lack of success in setting up an explant or reaggregate culture system that expands human AGM HSCs, although recent attempts have now succeeded in maintaining them in culture (Easterbrook *et al.*, 2019). Though the mouse model of developmental haematopoiesis is highly conserved with human and studies in mouse are very informative, it is important to study the human system to identify the clinically important human-specific traits of the first HSCs.

The Renin-Angiotensin pathway component Angiotensin-converting enzyme (ACE) has also been suggested to be a marker of HSCs in the AGM (Tavian, Hallais and Peault, 1999). Prior to the development of IAHCs, ACE marks scattered CD34-CD45- cells in the paraaortic splanchnopleure (P-Sp) as early as CS8-9 (19 – 21 dpc), which is the presumptive AGM. Between days 24 and 26 these CD34-CD45-ACE+ are able to generate haematopoietic colonies *in vitro* indicating a haematopoietic potential in the mesoderm prior to HSC generation in the dorsal aorta (Sinka *et al.*, 2012) which may indicate the presence of pre-HSCs. ACE then marks scattered cells in the mesenchyme around the dorsal aorta and from CS12 marks the endothelium and IAHCs (Jokubaitis *et al.*, 2008; Sinka *et al.*, 2012). ACE+ cells in the AGM have not been validated as HSCs with repopulation assays, however foetal liver HSCs at later stages (14 – 18 weeks tested) are entirely within the ACE+ fraction, and recently have been shown to mark highly proliferative HSCs in the foetal liver (Sinka *et al.*, 2012; Zhang *et al.*, 2019). However, ACE is also expressed in the YS in blood islands so it is not a specific marker of adult-type HSCs (Sinka *et al.*, 2012).

Not much is known about signalling within the human AGM niche and the pathways required for HSC formation, although components of the Notch, WNT and RA signalling pathways have been shown to be expressed on AGM endothelial cells and HSPCs (Ng *et al.*, 2016). A deeper understanding of the human HSC signalling niche will be of great use to the field and potentially provide key information that is missing from our studies with model organisms.

1.10 Generating HSCs in vitro – current progress

There is a huge amount of work in the field going into generating bona fide HSCs in culture, usually from pluripotent stem cells (PSCs). As discussed in chapter 1.3, the therapeutic implications of generating a patient-specific, unlimited source of fast and long-term repopulating HSCs are monumental for the treatment of numerous clinical disorders, namely malignancies of the blood system. Furthermore, in the same way that studies in developmental biology can influence HSC derivation culture conditions, the cultures can provide insights into the mechanisms behind HSC development *in vivo.* Many differentiation protocols for derivation of blood cells from PSCs (iPSCs/ ES cells) recapitulate changes that parallel *in vivo* haematopoietic development. There is therefore a loop of gaining understanding from the embryo and applying it to the culture system and vice versa which propels the field forwards. Thus far, bona fide HSCs have not yet been generated without transgenic

manipulations as depending on the protocol they lack one or more defining characteristics including, multipotency, long-term self-renewal and engraftment capacity (Wahlster and Daley, 2016).

1.10.1 Directed differentiation of HSC-like cells from pluripotent stem cells

The production of HSC/HSPC-like cells from PSCs typically involves recapitulation of the developmental progression through mesodermal followed by haematogenic endothelial intermediates (Chadwick *et al.*, 2003; Pick *et al.*, 2007; Ng *et al.*, 2016). As described in the preceding chapters, lessons from developmental biology show that HSC specification happens at multiple stages including both at the haematogenic endothelium and earlier in the LPM. A sequential protocol enables the fine-tuning of each intermediate cell type to optimise HSC generation. These intermediate steps are controlled by agonists and antagonists for signalling pathways in the embryo niche (Chadwick *et al.*, 2003; Pick *et al.*, 2007; Ng *et al.*, 2016). The mesoderm is specified using FGF, Wnt, BMP4, VEGF, SCF and activin signalling which, at this stage, must be patterned to differentiate between primitive and definitive haematopoiesis (Ng *et al.*, 2016; Wang and Nakayama, 2009; Nostro *et al.*, 2008; Kennedy *et al.*, 2012; Sturgeon *et al.*, 2014). This is a critical juncture as earlier protocols succeeded in generating only primitive-like haematopoieic progenitors (Ditadi, Sturgeon and Keller, 2016). Specification of the definitive program, as defined by T cell production, requires an increase in Wnt signalling and/or inhibition of activin and generates KDR+CD235a+ and KDR+CD235a- primitive and definitive-fated mesoderm respectively(Sturgeon *et al.*, 2014; Ng *et al.*, 2016; Kennedy *et al.*, 2012). This suggests that primitive and definitive waves of human haematopoiesis derive from subsets of cells from patterned mesoderm and not from the same progenitors which sequentially upregulate the primitive and definitive programs.

The haematogenic endothelium is then specified with additional cytokines and growth factors including FLT3L, TPO, IL6, IGF2 and FGF2 and is marked by the expression of SOX17. SOX17 expression is lost and RUNX1 is upregulated during the endothelial to haematopoietic transition in this culture system(Ng *et al.*, 2016). The HSPCs subsequently produced are considered to be more of a definitive type as they express HOXA genes similarly to repopulating HSCs from UCB samples(Ng *et al.*, 2016; Dou *et al.*, 2016). However, these cells do not provide long-term engraftment in irradiated NSG hosts and are therefore not true HSCs.

1.10.2 Reprogramming with transcription factors

An alternative approach to directed differentiation with cytokines and morphogens is to enforce expression of certain TFs in the PSCs. One study transfected human (h)ESC and iPSC derived CD34+ cells with *HOXA9, ERG, RORA, SOX4* and *MYB* and achieved generation of HSPCs with short-term engraftment and myeloerythroid output (Doulatov *et al.*, 2013). Follow up to this work led to the recent breakthrough whereby HSPCs with myeloid, B and T cell potential were capable of engrafting

both primary and secondary mice. The haematogenic endothelium again derived from hPSCs was transfected with a different combination of TFs; *ERG, HOXA5, HOXA9, HOXA10, LCOR, RUNX1* and *SPI1* determined through data mining of mouse and human foetal liver HSC datasets (Sugimura *et al.*, 2017). Alternative attempts were made to produce HSCs directly from somatic cell types bypassing the pluripotent intermediate. A group succeeded in generating HSC-like cells by direct reprogramming of human umbilical vein endothelial cells with the transcription factors *FOSB, GFI1, RUNX1* and *SPI1* and co-culture with vascular niche-like endothelial cells which generated long-term engrafting multi-potent cells but with limited T cell capacity (Sandler *et al.*, 2014). The same group have since achieved full multi-lineage reconstitution with a similar strategy using mouse endothelial cells but are yet to show this with human cells (Lis *et al.*, 2017).

These are exciting results but are not without their limitations. Some of the genes used for re-programming are oncogenes and may remain expressed in differentiated cells posing a heightened cancer risk (Hu *et al.*, 2011). Furthermore, random retroviral integration of the transgenes may result in malignant transformations (Guibentif and Göttgens, 2017). Although Lis et al., found no sign of cancer in mice engrafted with their mouse endothelial cell-derived HSCs (Lis *et al.*, 2017), a transgene free approach will be safer and more likely to move to clinic. Furthermore, the protocols are not very efficient as some recipient mice transplanted with reprogrammed cells were not repopulated. The HSCs produced with these methods were heterogenous and not yet identical to *in vivo* HSCs (Guibentif and Göttgens, 2017) so further efforts are required to generate sufficient quantities of reproducible HSCs. However, the promising results from these groups make the artificial generation of HSCs seem a feasible prospect. Lessons from development, particularly in the human embryo will help shape and refine these protocols to advance them to the clinic.

1.11 The rationale and goals for this thesis with an overview of the structure

This introduction provided an overview of the lessons learned from model organisms of the processes and signalling behind the spatiotemporally controlled development of adult-type HSCs in the embryo. The current understanding of HSC development in the human embryo was described and hopefully asserted that there is much to learn regarding the signalling behind human-specific HSC development. This incomplete picture has prohibited recapitulation of normal haematopoietic development in cultures of pluripotent stem cells and generation of bona fide adult-type HSCs with long-term multilineage engraftment. The aim of this thesis project was to generate a fuller picture of the signalling milieu in the human developmental HSC niche and in doing so, reveal potential key players which influence and promote HSC emergence.

This project utilises a multi-layered transcriptomics approach to generate a spatially resolved map of the gene expression and signalling pathways in the human AGM. First, laser capture microdissection (LCM) is coupled with RNA-Sequencing to generate a spatial transcriptome map of the dorsal aorta

and surrounding mesenchyme from human embryos within the window of HSC emergence. In short, this means that the positional identity of gene expression is not lost as with FACS sorted cells. This is particularly important as stem cells emerge in the context of the cells and signals around them and as signalling and HSC emergence in the AGM is polarised along the dorsal-ventral axis, retaining the positional information of niche signals is of particular importance. A spatial transcriptome also allows genes and pathways that are differentially expressed along the dorsal-ventral axis to be defined in a non-biased way and thus highlight putative signals that promote ventral HSC emergence. A spatial transcriptome approach was already shown to be useful in mouse with the elucidation of the BMP inhibitor BMPER as a novel regulator of HSC generation(McGarvey *et al.*, 2017).

Single cells and FACS sorted haemato-endothelial populations from the dorsal aorta were also sequenced and further resolve niche signals to cell types and states as well as enable modelling of the interactions between the niche and the haematogenic endothelium and nascent HSC/HSPCs. This approach, which allows transcriptome profiling of the dorsal aorta at the spatial, cell population and single cell level is key to having an in-depth analysis of the dynamic HSC developmental niche.

Furthermore, our analysis identified two candidates with a potential role in influencing HSC emergence, Renin (Ren) and Endothelin-1 (ET-1). With the addition of Endothelin-2 (ET-2), these factors were functionally shown to increase HSC production in mouse reaggregate culture. This highlights the value of the spatial transcriptome datasets obtained in this project. This is an important resource for the scientific community in uncovering novel players in the development of true, adult-type HSCs.

The data and analysis from this project are organised across 3 chapters with a discussion of the results at the end of each chapter and a final conclusion at the end. An overview of the results chapters (3-5) is as follows:

Chapter 3: A spatial transcriptome map of the human CS17 and CS16 HSC developmental niche

As LCM coupled with RNA-Sequencing (LCM-Seq) is not a very well-established technique and there are a number of technical difficulties, the first part of this chapter details the optimisation steps taken to establish it in a reproducible manner. This will hopefully inform and aid future users.

The second part of this chapter details the strategies used to sub-dissect, with LCM, the dorsal aorta and surrounding mesenchymal cells from CS17 and CS16 embryos into ventral and dorsal domains which were then sequenced. The differential gene expression (DGE) analysis and Gene Set Enrichment Analysis (GSEA) provides a comprehensive map of gene expression and signalling pathways that are polarised along the dorsal-ventral axis and at sites proximal and distal to the dorsal aorta and location of IAHCs. This chapter also identifies *REN* (Renin) and *EDN1* (Endothelin 1) as an example of ventrally enriched genes.

Chapter 4: Resolution of the spatial transcriptome through single cell and FACS sorted population sequencing.

In this chapter, endothelial, HSPC and haematopoietic populations were sorted from the ventral and dorsal domains of the dorsal aorta (CS15-16) and the transcriptomes were sequenced. This allowed elucidation of the dynamic changes in gene expression across the endothelial-haematopoietic transition (referred to as the EHT axis) as well as definition of an HSPC gene 'signature' with specific TFs, cell surface receptors and secreted factors. Furthermore, known interactions between niche (LCM-Seq) genes and EHT population genes were used to identify potential communication mechanisms between the niche and the haematogenic endothelium/HSPCs.

Single cell sequencing of the ventral dorsal aorta and niche cells explore in greater detail the dynamics and heterogeneity of the endothelium, identify potential haematogenic endothelium and important signalling populations. LCM-Seq niche signals are resolved to different cell types. Furthermore, renin and endothelin pathways components are resolved to distinct cell populations which facilitates modelling of how they may interact with the niche and haematopoietic cells. RNAScope and immunostaining also resolves gene and protein expression in the dorsal aorta revealing concentrated *EDN1* signalling below the IAHCs.

Chapter 5: Functional assessment of the role of Ren, ET-1 and ET-2 in promoting HSC emergence

Due to the scarcity of human tissue and the variability between samples, functional validation assays were carried out in mouse, which is a more tangible model for experimental analysis of HSC development. Gene expression of *Ren*, *Edn1* and *Edn2* were all found to be ventrally enriched in the mouse AGM as determined from a previously published dataset (McGarvey *et al.*, 2017) and the protein equivalents were added to a reaggregate and explant culture system from E9.5 and E10.5 respectively. Addition of each of these factors increased the complexity of haematopoietic colonies and, furthermore, was shown to increase HSC production compared to control as assayed by multilineage reconstitution of irradiated mice.

2. Materials and Methods

2.1 General Solutions

Dissection Buffer: Dulbecco's phosphate buffered saline (PBS) solution with Ca^{2+} and Mg^{2+} (Sigma-Aldrich), 4% foetal calf serum (FCS)(Gibco) and 50 units(U)/ml penicillin and streptomycin (P/S Gibco).

FACS buffer: PBS without Ca^{2+} and Mg^{2+}, 2% Foetal Calf Serum (FCS), and 50 U/ml P/S.

OP9 culture medium: Iscove's modified Dulbecco's medium (IMDM, Invitrogen), 20% FCS, L-glutamine (4 mM), P/S (50 U/ml).

Aggregate culture medium: IMDM (Invitrogen), 20% head-inactivated FCS (HyClone), L-glutamine/pyruvate (4mM), P/S (50 U/ml), 100ng/ml SCF, 100ng/ml IL3, 100ng/ml Flt3l (all Peprotech).

2.2 Human material

Human embryonic samples of Carnegie stages (CS) 15-17 were provided from two valuable resources, the MRC Centre for Reproductive Health and the MRC-Wellcome Trust Human Developmental Biology Resource. This study was approved by the Lothian Research Ethics Committee. The embryos were obtained immediately after elective termination of pregnancy for which each patient gave informed consent in writing. The embryos were dissected in dissection buffer. The embryos were dissected from the placenta and the heads removed for a different research group. The embryo bodies were either snap frozen in Optimal Cutting Temperature compound without prior fixation and stored at -80°C for later use in laser capture microdissection (LCM) and/or histology experiments or the embryo was dissected further. For experiments requiring dorsal-ventral (D-V) dissections of the dorsal aorta, the embryos were dissected in a similar manner to what has been described in the mouse (Medvinsky *et al.*, 2008). Generally, the body at the level of and including the heart was dissected away and discarded. Next, the neural tube was dissected off, followed by the ventral tissues including the gut and liver. Then, the ribs and limbs were dissected off followed by the mesonephros. Using the ventral branching vessels as a guide the dorsal aorta was sub-dissected into ventral (AoV) and dorsal (AoV) parts.

2.3 Animals

All mouse experiments were performed under a Project License granted by the Home Office (UK), University of Edinburgh Ethical Review Committee, and conducted in accordance with local guidelines. Animals were housed within the University of Edinburgh adhering to the Animals Scientific Procedures Act, UK, 1986. All mouse experiments were carried out by researchers with a personal licence granted by the Home Office. Mice were kept in stable light cycling conditions (14 hours light and 10 hours dark) with a regular supply of chow food and water. Embryos were used from C57BL/6 (CD45.2/2) (Jackson

Laboratories) mice and C57BL/6-Ly5.1 (CD45.1) mice were used as hosts for transplantation experiments. Flk1-GFP BAC transgenic mice were used in one LCM experiment (Ishitobi *et al.*, 2010).

To obtain embryos of the correct stage paired matings were set up and the morning of discovery of a vaginal plug is considered embryonic day (E) 0.5. The embryos were collected at E9.5 and E10.5. Following schedule 1 culling of the pregnant dam, the uterine horns were removed and the embryos dissected out of the extra-embryonic tissues including the yolk-sac and the amniotic sac. The embryos were then more accurately staged by counting the somite pairs (SPs) (E9.5 = 25 -29 SPs, E10.5 = 35 – 39 SPs). For E9.5 experiments the caudal parts were taken which includes everything below the heart. For E10.5 experiments the AGM region, comprising the dorsal aorta and mesonephros, were dissected as described for human.

2.4 Laser Capture Microdissection (LCM)

2.4.1 Cryosectioning

Before the experiment a cryotome FSE cryostat (Thermo Scientific) was sterilised with Leica Cryofect (Leica) and a paintbrush and microtome blade (MB22 Premier microtome blades (Thermo Scientific) were made nuclease-free using RNase AWAY (Thermo Scientific). Human and mouse embryonic samples frozen in OCT and stored at -80°C were equilibrated in the cryostat at a temperature of -24°C for one hour to prevent the tissue from being too brittle when sectioning. The embryo was sectioned, using the paintbrush to stop the tissue from rolling, in a caudal-to-rostral direction to a thickness of 7µm using MB22 Premier microtome blades (Thermo Scientific) and transferred to SuperFrost Plus slides (VWR). Frequent checks under the microscope verified the level reached along the rostral-caudal axis as defined by anatomical landmarks. Once the appropriate level had been reached the cryosections were transferred onto nuclease-free polyethylene napthalate (PEN) membrane slides (Zeiss). At intervals, a sister section would be transferred to a SuperFrost slide for future immunohistochemical analysis. During sectioning the PEN slides were always kept within the chamber at -24°C to prevent nuclease activity. To facilitate section transfer to the slide the region of the slide was warmed using a finger for 10 seconds. The completed slide was kept in the cryostat for > 5 mins and < 20 mins. The sections were then stained immediately.

2.4.2 Staining

Where possible nuclease-free reagents were acquired or aqueous solutions were made nuclease-free using diethyl pyrocarbonate (DEPC) (Sigma-Aldrich). Each step of the Cresyl Violet and Haematoxylin and Eosin (H&E) stains were carried out in sterile 50ml falcon tubes, 2 slides at a time, to minimise introduction of nucleases. For human LCM-Seq experiments the rapid H&E stains were used.

Cresyl Violet stain: PEN slides with tissue sections were fixed in ice-cold 70% ethanol for 3 min. The slides were then dipped in ice-cold DEPC treated H_2O for 1 min including 10 dips up and down to fully remove the OCT. The slides were dipped in cold 1% Cresyl Violet Acetate (dissolved in 50% ethanol, (Acros Organics)) for 30 sec. Additional 1 min incubations in cold 50% ethanol before and after the Cresyl Violet staining step were used in the enhanced protocol. The slides were then dipped in cold 70% ethanol followed by room temperature 100% ethanol for 1 min. The slides were air-dried within a PCR hood for 3 min then sealed in sterile falcon tubes and stored on dry ice for transfer to the LCM microscope.

Rapid Haematoxylin and Eosin stain: PEN slides with tissue sections were fixed in ice-cold 70% ethanol for 3 min. The slides were then dipped in ice-cold DEPC treated H_2O for 1 min including 10 dips up and down to fully remove the OCT. The staining steps were as follows; cold Mayer's Haematoxylin Solution (Sigma-Aldrich) for 4 min, cold DEPC-treated tap water for 2 min, cold Eosin Y (Sigma-Aldrich) 15 sec, cold 70% ethanol 1 min, cold 90% ethanol 1 min, room temperature 100% ethanol 3 mins. The slides were air-dried within a PCR hood for 3 min then sealed in sterile falcon tubes and stored on dry ice for transfer to the LCM microscope.

Immunostain: PEN slides with tissue sections were fixed in ice-cold 70% ethanol for 3 min. All staining steps were carried out on a staining tray sterilised with 70% ethanol and RNAse Away (Thermo Scientific). Sections were washed in PBS 3x 1 min then protein blocked in 1% Bovine Serum Albumin (BSA)/PBS (Sigma Aldrich) for 1 min. Sections were incubated with rat anti-mouse CD31 (BD Pharmingen, 0.2µg/ml) for 1 hour, washed 3x 1 min in PBS, incubated in secondary antibody goat anti-rat AlexaFluor (AF) 488 (Invitrogen, 10 µg/ml) for 1 hour and dehydrated in an ethanol series (75%, 90% and 100% ethanol) 30 sec each. The slides were air-dried within a PCR hood for 3 min then sealed in sterile falcon tubes and stored on dry ice for transfer to the LCM microscope.

Rapid Immunostain: PEN slides with tissue sections were fixed in ice-cold 70% ethanol for 3 min, protein blocked in 1% BSA for 1 min and washed in PBS 3x 1 min as before. Sections were incubated with a rat anti-mouse CD31-FITC (clone: 390, eBioscience, 5mg/ml) conjugated antibody at high concentration for 15 min. Finally, sections were dehydrated in an ethanol series (75%, 90% and 100% ethanol) and transferred to the LCM microscope straight away.

Alkaline phosphatase stain: PEN slides with tissue sections were fixed in 5% Dimethyl Sulfoxide (DMSO)(Sigma-Aldrich) in methanol (VWR) for 10 mins at room temperature. Sections were blocked in 1% BSA for 15 min and immunostained with anti-mouse CD31 (BD Pharmingen 0.5µg/ml**)**for 15 min, washed in 0.05% Triton X-100 (Sigma-Aldrich)/PBS (PBT) 3x 1 min, incubated with goat anti-rat IgG-AP (Cambridge Bioscience, 0.5 µg/ml) for 15 min, washed in PBT 3x 1 min, stained with the Vector® Blue Alkaline Phosphatase (Blue AP) Substrate Kit (Vector) according to the manufacturer's instructions. Finally, sections were washed 2x (30 sec) in PBS and dehydrated with an ethanol series (75%, 90%,

100%, 30 sec each). The slides were air-dried within a PCR hood for 3 min then sealed in sterile falcon tubes and stored on dry ice for transfer to the LCM microscope.

2.4.3 **LCM**

The LCM microscope used for the majority of the experiments was the PALM microbeam (Zeiss). The stained sections were transported to the LCM microscope sealed in falcon tubes on dry ice and brought to room temperature before opening the tubes to avoid condensation on the slides which enhances nuclease activity. The microscope and surrounding area were sprayed down with RNase AWAY to get rid of nucleases. Up to 3 slides were processed at a time and the time on the microscope stage was kept to under 40 minutes to minimise exposure to atmospheric nucleases. The sections were viewed and microdissected in brightfield using a 10X objective. The microdissected regions were collected into the caps of AdhesiveCap 500 opaque (Zeiss) 500µl PCR tubes. 15µl lysis buffer was added directly on top of the dissected tissue in the tube caps and the tubes closes in an inverted position – the lysis buffer used was determined by the RNA isolation method used (see next section). The inverted tubes were then kept on dry ice until all the sub-dissected regions had been collected into different tubes.

For the analysis of thick sections, 60 µm cryosections from Flk1-GFP transgenic mice were transferred onto PEN membranes mounted onto a steel frame (Leica). These were micro-dissected using a Leica LMD6 LCM fluorescent microscope which collects the sub-dissected regions by gravity.

2.4.4 **RNA purification, quantification and quality analysis.**

During the LCM-Seq optimisation different methods of RNA isolation and purification were trialled depending on the down-stream analysis. The RNeasy Micro Kit (Qiagen) was used to purify RNA from whole tissue sections and large LCM microdissected regions for subsequent RNA degradation analysis. The PicoPure RNA Isolation Kit (Thermo Fisher) was used to purify RNA from LCM microdissected samples for subsequent cDNA library preparation. Alternatively, the LCM samples were lysed using a mild hypotonic buffer (0.2% Triton X-100, 2U/µl RNase inhibitor, PBS (Nichterwitz *et al.*, 2016)) or lysis buffer from the CellsDirect One-Step qRT-PCR Kit (Thermo Fisher). The quality of the RNA was assessed using a TapeStation 2200 (Agilent) and High Sensitivity RNA reagents which provided an RNA Integrity Number (RIN^{e}). When concentrations of RNA were high enough, they were quantified using a Qubit fluorometer (Thermo Fisher Scientific) and Qubit RNA HS Assay Kit.

2.4.5 **Two-step PCR**

LCM lysates were amplified using the CellsDirect One-Step qRT-PCR Kit (Thermo Fisher). The housekeeping gene, murine TATA-binding protein (TBP) was amplified (Fwd: GGGGAGCTGTGATGTGAAGT, Rev: CCAGGAAATAATTCTGGCTCA). 25µl 2X Reaction Mix (CellsDirect, Invitrogen) and 0.2 µl RNase inhibitor (SUPERase-In Ambion) was added to 11.8 µl lysate/purified RNA.

1µl of Superscript III/Taq mix (CellsDirect) and 12µl of 10µM primers brought the reaction mix to 50µL. The thermal cycling conditions were 50°C for 15 min, 95°C for 2 min and 18 cycles of 95 °C for 15 secs and 60 °C for 4 min. Quantitative Real Time (qRT)-PCR was performed on triplicate dilutions of the cDNA using the Lightcycler 480 probes mastermix kit (Roche). Real Time PCR cycling conditions were as follows: 95°C for 5 min, 45 cycles of 95°C for 10s and 60°C for 18s, followed by 60°C for 10 min. For an RT-PCR positive control, whole E10.5 embryos were digested with collagenase as before and total RNA was extracted using the RNeasy Mini Kit (Qiagen). This was either diluted first to a 200 cell approximate concentration (57pg/µl) to be used as a pre-amplification control or SuperScriptVILO cDNA Synthesis (Invitrogen) was used for first strand synthesis of the undiluted RNA.

2.5 FACS-sorting human cells

Human dissected AoV and AoDs were dissociated to single cells in 1mg/ml Collagenase-Dispase (Roche) and 0.12 mg/ml of DNase I (Roche) for 35 min in a 37 °C rotating water bath. The cells were washed once in FACS buffer. Cells were immunostained with conjugated antibodies, anti-human Ve-Cadherin-PE (Beckman Coulter, 5µg/ml), anti-human CD45 –v450 (Clone: HI30, BioLegend, 6µg/ml) and anti-human CD235A – APC (BD Bioscience, 0.2µg/ml) for 1 hour at 4 °C. Cells were then washed twice and 7AAD was added. Cells were sorted using a FACS Aria-II sorter (BD Bioscience). Fluorescence minus one (FMO) staining was used to gate negative populations. Dead cells and erythroid cells were excluded by 7AAD and CD235A staining respectively. Sorted VC+CD45-, VC+CD45+ and VC-CD45+ populations were collected into Eppendorf tubes containing RLT buffer from the RNeasy Micro kit ready for RNA purification. Data acquisition and analysis was performed using FlowJo software (Tree Star).

For the 10X single cell sequencing, following AoV cell dissociation, CD34+ cells were isolated using the human CD34 MicroBead Kit and LS MACS Columns (both Miltenyi Biotec). This procedure was carried out as per the manufacturer's guidelines.

2.6 RNA-Seq Library Preparation

Two alternative RNA-Seq library preparation methods were used in this project. For CS17 LCM-Seq experiments and sorted population sequencing experiments the SMARTer® Stranded Total RNA-Seq Kit v2 -Pico Input Mammalian kit (Takara) was used. Input RNA was purified using the PicoPure RNA Isolation Kit and the protocol was followed according to the manufacturer's guidelines.

For CS16 LCM-Seq experiments a Smart-Seq2 protocol adapted for use with LCM was used (Picelli *et al.*, 2014; Nichterwitz *et al.*, 2016). 15µl hypotonic lysis buffer (See 2.4.4) was added on top of the LCM microdissected material in the caps of LCM collection tubes and pipetted up and down 5 times before closing the tube onto the lid in an inverted position. The cap side of the tube was vortexed for 15 seconds and spun in a tabletop centrifuge (8000g) for 5 min. 5µl lysate was added to 2µl 10mM dNTP

mix (Invitrogen) and 1µl 10µM oligodT (5'-AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3', (IDT)). This was vortexed briefly and spun in a microcentrifuge for 30 sec (700g) then incubated at 72°C for 3 min and immediately snap cooled on ice. To each reaction, 2 µl SSRTII 5 × buffer, 0.5 µl 100 mM DTT, 0.5 µl 200 U µl−1 SSRTII (all LifeTechnologies), 2 µl 5 M betaine (Sigma-Aldrich), 0.1 µl 1 M MgCl2 (Sigma-Aldrich), 0.25 µl 40 U µl−1 RNase inhibitor (Takara) and 0.1 µl 100 µM TSO-LNA-oligo 5'-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G -3', Exiqon) was added. The reverse transcription reaction was performed in a thermal heat cycler with the following conditions; 90min 42°C, 10 cycles of (2 min 50°C, 2 min 42°C) and 15 sec 70°C. For the amplification reaction 12.5 µl 2X KAPA HiFI Hotstart Mix (KAPA Biosystems), 0.2 µl 10 µM ISPCR primers (IDT, 5' - AAG CAG TGG TAT CAA CGC AGA GT – 3') and 2.3 µl nuclease-free H_2O (Invitrogen) was added to each reaction. This was heat cycled as follows: 3 min 98°C, 18 cycles of (20 sec 98°C, 15 sec 67°C, 6 min 72°C) and 5 min 72°C. After bead purification using AMPure XP beads (Beckman Coulter), the concentration of the cDNA library was measured with an Agilent 2200 TapeStation using the High Sensitvity DNA 5000 kit (Agilent). 1 ng of cDNA from this reaction was amplified and barcoded using the Nextera XT DNA sample preparation kit and Nextera XT index kit (Illumina) following the manufacturer's protocol. The libraries were purified again using AMPure XP beads, analysed on the Tapestation 2200 using High Sensitivity DNA 100 Kit and quantified using the Qubit fluorometer and Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific).

2.7 RNA Sequencing

LCM-Seq and bulk-sorted population RNA sequencing was carried out at Edinburgh Genomics on a NovaSeq SP flowcell generating 50 base pair (bp) or 75 bp paired end reads. LCM-Seq samples were sequenced at a read depth of approximately 29 million reads per sample. Bulk-sorted population samples were sequenced at a read depth of approximately 47 million reads per sample. 10X samples were sequenced on a NovaSeq S1 flow cell with a 26/8/91 cycle set up at a read depth of approximately 70,000 reads per cell.

2.8 Read quality control, alignment and counting

The quality of the reads including Phred quality score (quality of the identification of nucleobases during sequencing) adapter contamination, GC content, duplicate levels was assessed using the tool FastQC (Andrews, 2010). Illumina adapter sequences were trimmed off reads using the tool Flexbar (Dodt *et al.*, 2012). Reads were mapped to the human reference genome hg38 (Ensembl version 85) using STAR (Dobin *et al.*, 2013). SAMtools (Li *et al.*, 2009) was used to sort and index the aligned reads. The multicov function of BEDtools (Quinlan and Hall, 2010) was used to count the read fragments per gene and generate a matrix of reads per gene.

2.9 Differential gene expression (DGE) analysis

In R, the tool DESeq2 (Love, Huber and Anders, 2014) was used for DGE analysis which provides $\log_2$ fold change of genes between two conditions. Per embryo batch effects were corrected for using the following design: ~embryo + population. The p. value is dependent on the statistical test used within the tool which was either the Wald test or the likelihood ratio test (LRT) as indicated throughout the results chapter. P values were adjusted for by multiple hypothesis correction with the Benjamini and Hochberg method to produce adjusted p. values (p.adj)(Benjamini and Hochberg, 1995). Genes were considered significant with a p.adj < 0.05. Principle component analyses were carried out on variance stabilizing transformed datasets using in-built R statistical tools and visualised using ggplot2 (Wickham, 2019). Heatmaps were also carried out on variance stabilizing transformed datasets using in-built R statistical tools and visualised using the R package pheatmap. Volcano plots were generated using EnhancedVolcano (Blighe, 2018).

2.10 Pair-wise Gene Set Enrichment Analysis

Enriched gene sets were found from ranked lists of differentially expressed (DE) genes between two conditions using the GSEA tool from the broad institute (Subramanian *et al.*, 2005). DE genes lists were ranked by log2 fold change. Annotated gene sets from the following databases were used; Hallmark gene sets (Liberzon *et al.*, 2015), BioCarta (Nishimura, 2004), KEGG (Kanehisa and Goto, 2000), Reactome (Croft *et al.*, 2011) and Gene ontology (GO) (Ashburner *et al.*, 2000; 'The Gene Ontology Resource: 20 years and still GOing strong,' 2019). Genes were input as a ranked list and pathways were considered significant with a false discover rate (FDR) of <0.25.

2.11 Finding protein association networks

The String-DB was used to find interactions between proteins from lists of genes (Szklarczyk *et al.*, 2015; Szklarczyk *et al.*, 2019). These interactions are both direct (physical) and indirect (functional) interactions. The tool determined interactions between proteins by known interactions from curated databases or that had been experimentally determined, predicted interactions from gene neighbourhood, gene fusions and gene co-occurrence databases, and by text mining publications, co-expression and protein homology databases. Interactions were given a confidence score scaled between 0 and 1 which is a combined score for each type of evidence given and is the estimated likelihood that a given interaction is 'biologically meaning, specific and reproducible'. Only high confidence score (0.7) interactions were used. The network matrices generated in the string-db web interface were opened in the Cytoscape software (Shannon *et al.*, 2003) for production of figures and scaling of nodes by the number of interactions associated with it.

2.12 10x Single Cell Sequencing

Libraries from 10,000 CD34+ AoV cells were prepared using the Single Cell 3' Reagent Kit v2 (10x Genomics) across two lanes of the Chromium™ Single Cell Chip (10x Genomics). The libraries were prepared according to the manufacturer's protocol.

2.13 10x Single Cell Analysis

The Cell Ranger 2.1.0 (10x Genomics) analysis pipeline was used to process the 10x single cell RNA-Seq output by aligning reads to GRCh38 human transcriptome (Ensembl), generating feature-barcode matrices and performing t-SNE clustering. The single-cell gene expression data were converted into a loom file using the python tool Loompy (https://github.com/linnarsson-lab/loompy). SCANPY (Wolf, Angerer and Theis, 2018) was used for the follow steps 1 – 13. 1) Filtering out cells with less than 200 genes. 2) filtering out genes that were in less than 3 cells. 3) Filtering out cells with a) high percentage mitochondrial genes (>0.05) and b) a high number of genes (> 3500). 4) Normalising reads per cell to 70,000. 5) Logarithmise the data. 6) Regress out the effects of total counts per cell and the percentage mitochondrial genes. 7) Regress out cell cycle effect. 8) Principal component analysis. 9) Compute the neighbourhood graph (n_neighbours = 18). 10) Embed the neighbourhood graph in 2 dimensions using Uniform Manifold Approximation and Projection (UMAP) (McInnes, Healy and Melville, 2018). 11) Clustering the neighbourhood graph using Louvain method (Levine *et al.*, 2015). 12) Finding marker genes per cluster using a Wilcoxon rank-sum (Mann-Whitney-U) test. 13) Mapping gene expression onto the UMAP embedding. Partition-based graph abstraction (PAGA) (Wolf *et al.*, 2019) was used to make lineage inferences from the neighbourhood graph, using the Louvain clustering. scVelo (https://github.com/theislab/scvelo) was used to estimate the direction of RNA Velocity (Manno *et al.*, 2018) across the UMAP embedding and in turn infer lineage trajectories from the PAGA-inferred cluster relationships.

2.14 Immunohistochemistry

SuperFrost Plus (VWR) slides with 7µm cryosections were taken out of -20 °C storage and placed immediately in cold 4% PFA (Sigma Aldrich) for 10 min. Staining steps were as follows; 3x wash in PBS, 5 min each. 10 min permeabilisation in PBS/0.5% Triton-X100 (Sigma Aldrich). 2x 5 min PBS wash. 30 min PBS/10% FCS protein block. Overnight incubation with a primary antibody diluted in PBS/2% FCS. 2x 5 min PBS wash. 2 hour incubation with secondary antibody diluted in PBS/2% FCS at room temperature. 2x 5 min PBS wash, 5 min 30nM DAPI (ThermoFisher) incubation. 1x 5 min PBS wash.

Mount in Prolong Gold Antifade (ThermoFisher) and coverslip of thickness 1.5 (VWR). Slides were left at room temperature in the dark overnight before imaging. Anti-human primary antibodies are described in Table 1. Anti-mouse primary antibodies are described in Table 2. Secondary antibodies are described in Table 3. Imaging was done using a 5 laser Confocal TCS SP8 (Leica) and an Axio Observer Z1 Inverted Microscope (Zeiss). For each staining experiment there was a negative control with no primary antibodies and all secondary antibodies or with an isotype control and all secondary antibodies.

Antigen	Clone	Isotype	Conjugate	Supplier	Working conc.
VE-Cadherin		Mouse	NA	BD Pharmingen	1µg/ml
CD31		Sheep	NA	R&D	2µg/ml
Runx1	EPR3099	Rabbit	NA	Abcam	1.6 µg/ml
CD45		Mouse	NA	BD Pharmingen	2 µg/ml
Renin		sheep	NA	Abcam	4 µg/ml
Endothelin-1	TR.ET.48	Mouse	NA	Abcam	10 µg/ml
Endothelin Receptor A	485711	Mouse	NA	R&D	5 µg/ml
Endothelin Receptor B		Rabbit	NA	Abcam	10 µg/ml

Table 1. Human primary antibodies

Antigen	Clone	Isotype	Conjugate	Supplier	Working conc.
CD31	MEC13.3	Rat	NA	R&D	2 µg/ml
cKit		Goat	NA	R&D	5 µg/ml
Endothelin-1		Rabbit	NA	abcam	8 µg/ml

Table 2 Mouse primary antibodies

Antigen	Clone	Isotype	Conjugate	Supplier	Working conc.
sheep IgG		Donkey	NL557	R&D	2 µg/ml
mouse IgG		Goat	AF488	Invitrogen	2 µg/ml
rabbit IgG		Donkey	AF647	abcam	1.9 µg/ml
Rat IgG		Donkey	AF546	Life tech.	2 µg/ml
Goat IgG		Donkey	AF488	Invitrogen	2 µg/ml

Table 3 Conjugated secondary antibodies

2.15 RNAScope

RNAScope was carried out using the RNAscope® Multiplex Fluorescent Reagent Kit v2 (bio-techne) largely following the manufacturer's instructions for use with fresh frozen samples. Differences were an initial 30 min fixation of the 7µm cryosections in 4% PFA at 4°C and 30 min incubation at 37 °C following dehydration. Protease IV was applied for 10 min at RT. The fluorophores used were all Opal™ Dyes (Perkin Elmer); Opal 520, Opal 570 and Opal 690. The probes used were all human RNAscope® Probes; Hs-EDN1, Hs-EDNRA-C2, Hs-EDNRB-C3, Hs-REN-C3. Stained slides were imaged using a 5 laser

Confocal TCS SP8 (Leica) and an Axio Observer Z1 Inverted Microscope (Zeiss). Positive and negative control slides were also stained using positive and negative probes supplied in the kit.

2.16 Image analysis

All image analysis was done using ImageJ FIJI software (Schindelin *et al.*, 2012). To quantify RNAScope *EDN1* hotspots (clusters of probe signals) the image threshold was first adjusted to filter out weak and background signals and a binary colour image of probe signal vs. non-signal was generated. Process → Binary → Watershed was used to help separate merged signals. The analyse tool was then used to highlight any signal clusters that were 300 pixels or larger.

2.17 Ex vivo floating membrane aggregate and explant cultures

E9.5 caudal parts from CD45.2/.2 BL6J mice were dissociated in 1mg/ml Collagenase-Dispase (Roche) and 0.12 mg/ml of DNase I (Roche) for 35 min in a 37 °C rotating water bath. FACS buffer was added to neutralise the collagenase and cells were centrifuged for 5 min at 300g. Supernatant was aspirated and tissues mechanically dissociated by gentle pipetting up and down in FACS buffer. Cells were then centrifuged and washed again. Cells were then resuspended in IMDM/20% heat inactivated FCS (Hyclone) with addition of one of the recombinant proteins; Endothelin-1 (LS-Bioscience), Endothelin-2 (Sigma Aldrich), Renin (Sigma Aldrich) at a concentration of 10ng/ml or 100ng/ml. A no-protein control was also included. Dissociated cells were distributed into p200 pipette tips sealed at the ends with paraffin; 1 embryo equivalent (ee) per tube in 20µl. Tips were centrifuged at 460g for 12 min to form a pellet. The pellets were carefully deposited onto 0.8µm nitrocellulose membrane filters (Millipore) floating on Aggregate Media (control or + 1 recombinant protein at 10ng/ml or 100ng/ml). A maximum of 5 aggregates were cultured per membrane at the gas-liquid interface for 7 days at 37°C, 5% CO_2. E10.5 AGMs were cultured as whole explants on floating membranes as before. After 7 days culture the colonies or explants were dissociated into single cells in 1mg/ml Collagenase-Dispase (Roche) and 0.12 mg/ml of DNase I (Roche) for 35 min in a 37 °C rotating water bath.

2.18 Haematopoietic colony-forming unit assays

After 7 days reaggregate or explant culture, collagenase dissociated cells were plated at a concentration of 0.05 and 0.01 ee of the starting culture in Methocult™ GF M3434 (StemCell Technologies Inc.). Plates were incubated at 37°C, 5% CO_2 for 7 days. Haematopoietic colonies were counted and scored after 7-11 days of differentiation following described standard criteria (Pereira, Clarke and Damen, 2007). For this study, Colony-forming unit-macrophage (CFU-Mac), colony-forming unit-granulocyte, macrophage (CFU-GM), colony-forming unit-granulocyte, erythroid, macrophage, megakaryocyte (CFU-GEMM) were counted.

2.19 Long-term repopulation assay

Collagenase dissociated cells from E9.5 7 day aggregate cultures were washed and resuspended in FACS buffer with heat-inactivated FCS (hyclone). CD45.1/.2 mice were irradiated prior to transplantation with a dose of 9.5 Gy, split in two doses with a 3-hour gap in between and delivered by a sealed Cs source at a rate of 21.6 rad/min. 0.5ee were injected with a 30-gauge syringe needle into the lateral tail veins of pre-irradiated CD45.1/.2 mice with 100,000 bone marrow carrier cells isolated from the femurs of CD45.1/.1 mice. Blood was collected from the tail vein by superficial incision of the tail vein at 8 and 18 weeks post-transplantation to analyse blood chimerism. At 18 weeks the mice were culled and haematopoietic tissues taken for multilineage analysis including the spleen, bone marrow, thymus and blood. Bone marrow was flushed from the femur with FACS buffer using a syringe and 30-gauge needle. Thymus and spleen were chopped with dissection scissors into fragments and pipetted up and down to flush out the haematopoietic cells before filtering out the remaining tissue. Red blood cells in the blood were lysed with Red Blood Cell Lysis Buffer (bioRad) for 15 min at room temperature. Cells were then stained as described below.

2.20 Flow cytometric analysis of cells

After 7 days reaggregate or explant cultures collagenase dissociated cells were immunophenotypically analysed by flow cytometry on a BD LSR Fortessa (BD). Cells were washed in FACS buffer and centrifuged for 5 min at 300g. Two panels of antibodies were used to stain cells shown in Table 4 and Table 5. Cells were incubated with the antibodies for 1 hour at 4°C, centrifuged and washed as before then stained with BV711 Streptavidin (BD Horizon) for 1 hour at 4°C. Cells were centrifuged and washed again. 1µM 7AAD (Invitrogen) was added to the cell suspensions before flow cytometric analysis.

Antigen	Clone	Isotype	Conjugate	Supplier	Working conc
CD41	MWReg30	Rat	BUV379	BD Horizon	0.6µg/ml
CD44	IM7	Rat	bv421	BioLegend	1 µg/ml
cKit	ACK2	Rat	bv510	BioLegend	1 µg/ml
CD45	30-F11	Rat	V650	BD Horizon	2 µg/ml
EPCAM	G8.8	Rat	biotin	eBioscience	2 µg/ml
CD146	ME-SF1	Rat	PE	BioLegend	1 µg/ml
Sca-1	AVD7	Rat	PEcy7	BD Pharmingen	2 µg/ml
VE-Cadherin	eBioBV13	Rat	e660	Invitrogen	2 µg/ml

Table 4: Panel 1 antibodies for flow cytometry analysis of haematopoietic cultures

For the assessment of donor contribution in reconstituted mice at 8 weeks and 18 weeks post-transplantaion of E9.5 aggregates, rat anti-mouse CD45.1-APC (Invitrogen, 2 µg/ml, Clone: A20) and

rat anti-mouse CD45.1-PE (eBioscience, 2 μg/ml, Clone: 104) were used to stain the peripheral blood cells.

Antigen	Clone	Isotype	Conjugate	Supplier	Working conc.
CD45	30-F11	Rat	BUV379	BD Horizon	2 µg/ml
CD41	MWReg 30	Rat	bv421	BioLegend	0.6 µg/ml
Ter119	TER-119	Rat	bv510	BD Horizon	2 µg/ml
CD11b	M1/70	Rat	V500	BioLegend	0.03 µg/ml
b220	RA3-6B2	Rat	V650	BD Pharmingen	2 µg/ml
CD3	145-2C11	Rat	biotin		2 µg/ml
Gr1	RB6-8L5	Rat	PE		1 µg/ml
CD11c	N418	Rat	PEcy7	BioLegend	2 µg/ml
CD335	29A1.4	Rat	e660		4 µg/ml

Table 5: Panel 2 antibodies for flow cytometry analysis of haematopoietic cultures

Antigen	Clone	Isotype	Conjugate	Supplier
CD45_1	A20	Rat	V450	BD Horizon
CD45_2	104	Rat	V500	BD Horizon
b220	RA3-6B2	Rat	APC-cy7	eBioscience
CD11c	N418	Rat	PE-cy7	BioLegend
Ter119	TER-119	Rat	FITC	eBioscience
Gr1	RB6-8C5	Rat	PE	eBioscience
CD335	29A1.4	Rat	BV711	BioLegend
CD3e	145-2C11	Rat	APC	eBioscience
CD4	GK1.5	Rat	APC	eBioscience
CD8	53.67	Rat	BV711	eBioscience

Table 6: Antibodies used for multilineage analysis of mice with donor reconstitution

For the multi-lineage analysis 18 weeks after transplantation of E9.5 aggregate cultures, cell suspensions from the thymus, spleen, bone marrow and blood were stained with antibodies as above. The antibodies used are found in Table 6. In all cases, FMOs were used to gate negative populations. OneComp and UltraComp beads (both ThermoFisher) were stained with single antibodies for automatic compensation by the Fortessa. All data were analysed on FLOWJo software (BD).

3. Spatial transcriptome mapping of the human AGM

In order to map the spatial transcriptome the technique of laser capture microdissection (LCM) coupled with RNA-Seq (LCM-Seq) was employed. The Zeiss PALM LCM microscope uses a laser to finely dissect out regions of interest from tissue sections on a histology slide. Using a pulse of energy, it then catapults the dissected region into the cap of a collection tube directly above. By adding lysis buffer, you can then extract the RNA from the tissue and, from this, generate libraries for RNA-Sequencing. Differential gene expression analysis (DGEA) between different sub-dissected regions allows you to define gene expression patterns and molecular gradients across the region of interest.

In this study, LCM-Seq will explore patterns of gene expression in the human HSC developmental niche, the AGM region. The tool is used to assess dorsal-ventral (D-V) differences in signalling within the AGM region, both at close range to the dorsal aorta and more distant, broader signalling patterns throughout this domain. The goal was to sub-dissect the endothelial layer into domains along the D-V axis and, in later experiments, additional sub-dissections of the underlying perivascular and mesenchymal cells.

In setting up the technique, there were a number of challenges to overcome:

1) RNA is highly susceptible to degradation
    I. The embryo must be sectioned on to slides, stained for visualisation of the cells, transferred to the LCM-Microscope and exposed on the microscope stage during microdissection – all without too much RNA degradation.
    II. Histological staining of the sections must therefore be adequate to identify cells by morphology or immunohistochemistry without severely impacting RNA quality.
2) Quantities of RNA are very low
    I. Sufficient RNA must be extracted from the dissected lysed tissues to prepare RNA-Seq libraries.

3.1 **LCM-Seq optimisation**

The LCM-Seq technique was optimised using E11.5 mouse embryos snap frozen in OCT without fixation in PFA as formaldehyde cross-links with macromolecules in the tissue reducing the yield of extracted RNA (Evers *et al.*, 2011). These embryos were then sectioned onto PEN slides (specialised nuclease-free slides with a membrane to facilitate LCM) in a cryostat sterilised using low temperature disinfectant and with blade, brush and slides made nuclease free with RNAse Away. The tissue sections were always kept within the cryostat at -24°C and stained within 20 minutes of being sectioned.

A critical optimisation step was to stain the tissue sections to achieve visualisation of the endothelial cells of the dorsal aorta and other structural components of the AGM without impacting the RNA

quality too severely. The membrane on the slides for LCM worsens the cell morphology under the microscope and the lack of mountant or coverslip makes high quality imaging particularly difficult to attain.

3.1.1 **Histology staining optimisation for LCM**

Four different staining methods were tested for their suitability for LCM-Seq. The precise details of each method are in the 'Materials and Methods' section 2.4.2. Each method was streamlined to be as quick as possible, and reagents were fresh, non-contaminated and DEPC-treated where possible. The impacts of the staining protocol were assessed by extracting RNA from freshly sectioned, stained or unstained tissue sections using the RNeasy Micro Kit. The purified RNA was then run on a TapeStation 2200 to get an RNA Integrity Score (RIN^e) out of 10 based on the ratio of 18S and 28S ribosomal RNA (rRNA) peaks on the electropherogram as well as the appearance of degraded products.

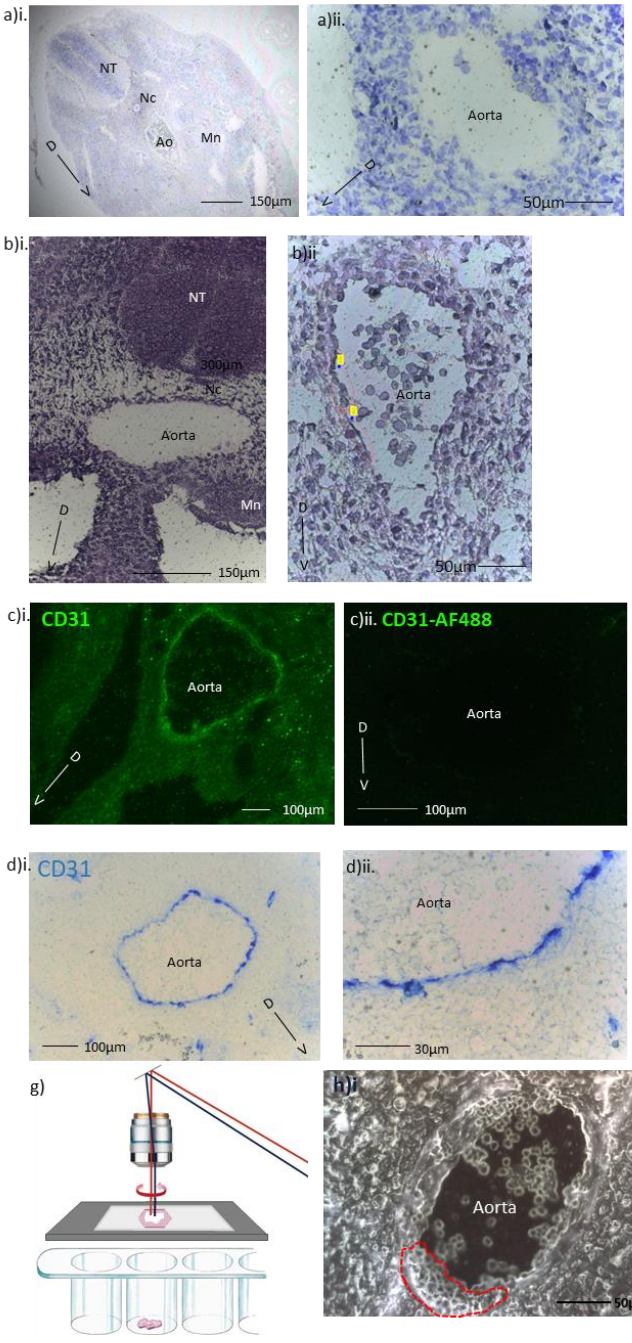
Cresyl Violet (CV) staining is recommended by Zeiss for use in LCM with subsequent RNA-extraction as all the solutions contain high ethanol concentrations which minimise RNA degradation. The stained tissue sections showed an average reduction in RIN^e of only 0.5 (n=3) from the RIN^e of the non-stained control (Fig.8e). However, the morphology of the cells was very poor when the tissue sections were stained on the LCM membrane slides with indistinct pale staining and no definition of the endothelial layer and IAHCs (Fig.8ai). Increasing the time of CV staining from 30 seconds to 2minutes and/or reinforcement with 50% Ethanol pre and post CV as suggested in the Zeiss Manual (Fig.8aii) which, although made the stain darker, did not greatly improve visualisation of cell morphology.

A Haematoxylin and Eosin stain was also trialled. This gave an average reduction in RIN^e of 0.3 (n=3) (Fig.8e) and furthermore, greatly improved the cell morphology and resolution of the anatomical structures in the surrounding tissues (Fig.8bi, ii). However, although visualisation of the dorsal aorta endothelial layer was enhanced compared to the CV stain it was not clearly defined.

An immunostain with a CD31 antibody and AF488 secondary antibody was used to mark the endothelial layer which did improve visualisation of the endothelial layer (Fig.8c.i) but caused a severe reduction in RIN^e of 6.2 (n=3), which may in part be due to the 2 hour total staining time compared with 15 mins and 12 mins for CV and H&E stains respectively. A more rapid immunostain was devised with a conjugated CD31-AF488 antibody which takes 19 mins in total. This in fact, resulted in an even larger decrease in RIN^e (7.6, n=1). By measuring the RIN^e of a tissue section after each step in the staining protocol it appeared that the blocking step in 1% BSA gave the largest RIN^e reduction of 5.5 (Fig.8f) Removing this step may have improved the RIN^e but as the signal strength from this method was very weak, it wasn't pursued further (Fig.8cii). Another drawback of using fluorescence was that the LCM microscope could only dissect in the brightfield which meant locating the cells under fluorescence then switching back to brightfield to dissect which added significant time to the

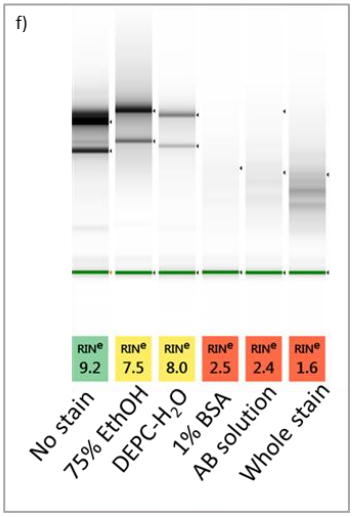
dissection of multiple tissue sections. This time needs to be kept to a minimum to reduce RNA exposure to nucleases.

The fourth stain was an alkaline phosphatase stain against CD31 antibody. This showed the endothelial cells well under the brightfield setting, however all other cell morphology and tissue structures were lost (Fig.8di, ii). Furthermore, the reduction in RIN^e after staining was high at 5.8 (Fig.8e). An H&E stain was thus considered to be the most suitable for LCM-Seq as it provided decent cell morphology and a very low ΔRIN^e.



e)

Staining Method	ΔRIN^e
Cresyl Violet	0.5
Haematoxylin and Eosin	0.3
Immunostain	6.2
Rapid Immunostain	7.6
Alkaline-Phosphatase Immunostain	5.8



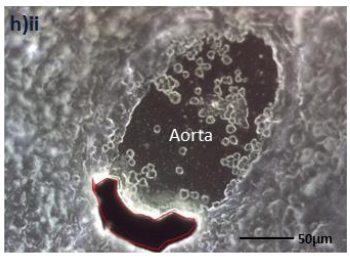


Figure 8. Optimisation of LCM staining protocols. (a-d) are representative images of individual staining protocols trialled for LCM-Seq; (ai) Cresyl Violet and (aii) Cresyl Violet enhanced stain, NT = neural tube. Ao = Dorsal Aorta. Nc = Notochord. Mn = Mesonephros (b) Rapid Haematoxylin and Eosin stain, (ci) anti-CD31 + AF488 Immunostain and (cii) CD31-AF488 conjugate rapid Immunostain, (d) Alkaline phosphatase anti-CD31 Immunostain. (e) shows the amount of change (Δ) in RIN^e for each staining protocol compared to the RIN^e of a sister section without any staining. (f) shows the RIN^e for each step in the rapid Immunostain, EthOH = Ethanol, AB = antibody. (g) Adapted from: https://www.leica-microsystems.com/products/light-microscopes/p/leica-lmd7/. It shows the leica system of microdissection with collection of the sub-domains by gravity into a tube below. (h) shows representative images of 60µm Flk-1-GFP cryosections before h)i and after h)ii laser microdissection with the leica system.

3.1.2 **Optimisation of low quantity RNA extraction**

The quantities of RNA from microdissected samples is in the picogram range and is below the threshold of reliable detection of most RNA quantification equipment. It is therefore difficult to test whether there are sufficient quantities of quality RNA in microdissected samples for RNA-Seq. Real-Time PCR (RT-PCR) was used to verify RNA extraction from microdissected samples and thus to optimise the protocol.

A two-step RT-PCR was carried out with initial cell lysis, cDNA synthesis of total RNA and pre-amplification steps using the CellsDirect one step qRT PCR Kit (Invitrogen) followed by a UPL real-time PCR assay of the house-keeping gene TATA-binding protein (TBP). Quantitative real time(qRT)-PCR was carried out on LCM microdissected dorsal aorta (Ao) endothelial cells from 7µm E11.5 CV stained cryosections. To test whether successful RNA extraction is dependent on the size of area microdissected, samples of approximately 200 (n=3) and 300 cells (n=3) (about 30000µm^2 and 45000µm^2 respectively) were dissected as well as large rectangular dissected areas (~150,000µm^2) (n=3). A pre-amplification control was used including RNA extracted from a whole E10.5 embryo then diluted to 57pg/µl, a concentration estimated for 200 cells, to confirm that the initial pre-amplification step worked. Undiluted whole embryo RNA was used as a second step RT-PCR control for which the first strand cDNA synthesis was carried out but not pre-amplification (Ctrl +). Finally, there was also a control without the initial reverse transcription reaction, to check that contaminating DNA wasn't being amplified.

No amplification was observed in LCM samples 1 and 2 from 200 and 300 cells respectively (Fig.9a). There was amplification seen however in the pre-amplification control and the RT-PCR control (Ctrl +) indicating that both steps of the RT-PCR are working correctly. Sample 3, from a large microdissected area, showed amplification for 1 in 10 and 1 in 100 dilution factors. Despite no RNA detection in the undiluted sample this indicates that from larger amounts of microdissected starting material RNA, there is sufficient RNA to enable target gene amplification. The cycle threshold value (Ct), which is the number of PCR cycles required to reach a signal threshold at the beginning of the exponential phase, was quite high however, with an average of 36.3, indicating a low amount of starting material. Prior to LCM the RIN^e of whole tissue sections were assessed following CV staining and gave a high value of 9.2. This indicates that either the RNA has degraded dramatically during LCM that time or that the method of RNA extraction from the cells is inefficient.

Therefore, four methods of RNA extraction were trialled side by side including the RNeasy Micro Kit (Qiagen), the PicoPure RNA Isolation Kit (Thermo Fisher), direct lysis using a mild hypotonic buffer and RNase Inhibitor and the method used previously – direct lysis and amplification using the CellsDirect™ kit (Thermo Fisher). 50,000µm^2 (around 350 cells) were microdissected from 7µm E11.5 CV stained cryosections, RNA was extracted using one of the four methods and a two-step RT-PCR for TBP was

carried out as previously. Three replicates were run for each condition with three PCR technical replicates for each condition replicate. The direct lysis method using the CellsDirect kit this time had an amplification product but had the highest mean Ct value, 33, (Fig.9b) indicating the lowest initial yield of RNA, followed by the RNeasy Micro Kit at 31.8. Direct lysis using a mild hypotonic buffer had a lower mean Ct of 30 and the PicoPure RNA Isolation Kit had the lowest mean Ct of 29.5. These two latter RNA extraction methods were used in subsequent RNA-Seq library preparation depending on the method of library preparation used.

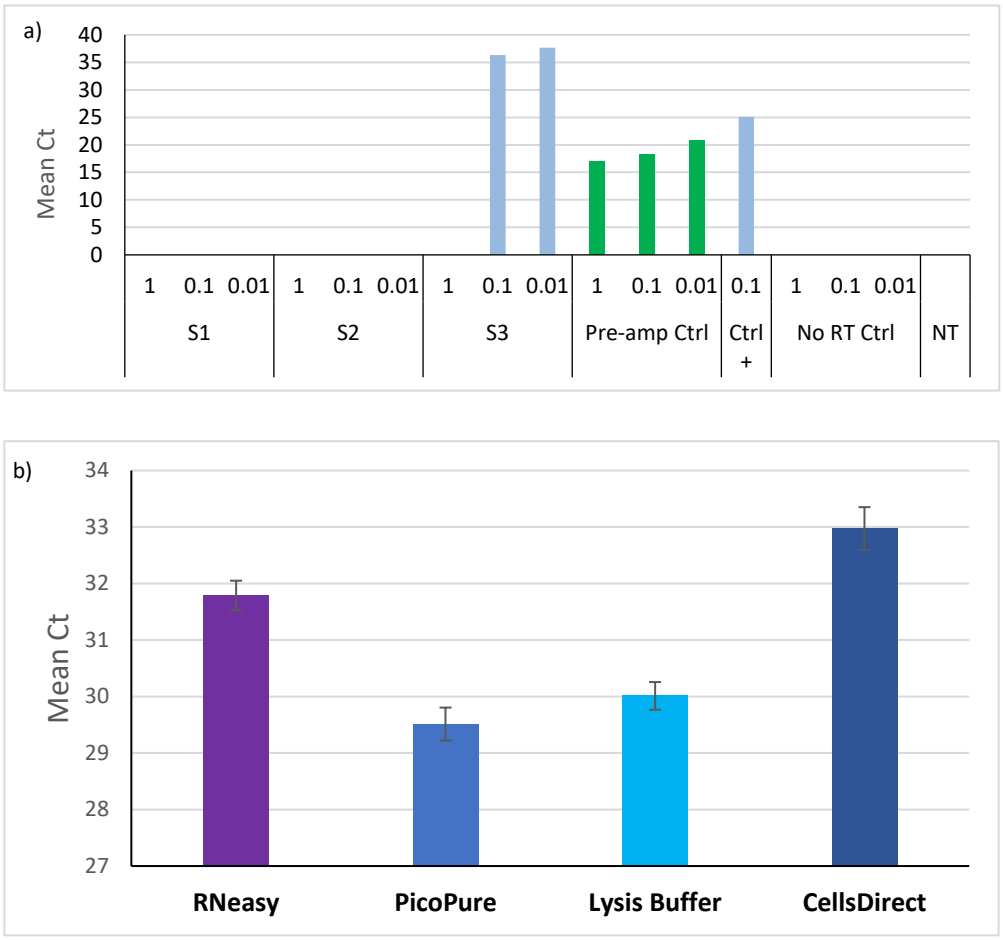


Figure 9 qRT-PCR analysis of RNA from E10.5 LCM microdissected cryosections. a) Mean cycle threshold. (Ct) values for 2-step RT-PCR of LCM cells. S1 = ~200 cells, S2 = ~300 cells, S3 = ~1000 cells in 3 dilutions; no dilution (1), 1/10 dilution (0.1) and 1/100 dilution (0.01). NT = No template control (H_2O). N=1. (b) Mean Ct values for 2-step RT-PCR of ~350 LCM cells after RNA extraction by different methods. Error bars show standard deviation around the mean, N=3.

3.1.3 Trial with thick sections and LCM with gravitational collection

To address the issue of low cell numbers and RNA yield obtained using Zeiss PALM LCM from thin (7µm) sections another strategy was devised using thicker sections of 60µm. The catapulting technology of the PALM microscope does not work for such thick sections so another laser capture

microscope from Leica was trialled (Leica LMD6). In this system the tissue is placed on top of a membrane in a frame. The laser microdissects around the tissue and the membrane and it is collected by gravity into a collection tube below (Fig.8g). Flk1-GFP embryos were used to visualise the endothelium of the dorsal aorta. Flk1 is a receptor to the angiogenic vascular endothelial growth factor (VEGF)-A and is essential for blood vessel development and organisation (Shalaby *et al.*, 1995).

60µm cryosections were placed onto the framed membranes and visualised under the microscope. The GFP signal of the tissue section was indistinct, possibly due to high background autofluoresence from the tissue which retained moisture (Fig.8hi-ii) but the endothelium could still be visualised. The main issue with this technique was that the endothelial lining was not exactly perpendicular to the membrane and lies at an angle which can be visualised in Fig.8hi-ii. As the laser cannot cut at an angle but only directly down, it is difficult to define exactly what has been dissected beneath the top layers of cells. This problem is much less exaggerated with thin sections. For these reasons this technique was abandoned.

3.1.4 **Optimisation of LCM-Seq library preparation with human embryonic samples**

Due to the sensitivity and rarity of early stage human embryo samples along with the high cost of RNA-Seq library preparation kits it was not appropriate to lose resources optimising the human LCM and RNA-Seq library preparation. The tools and techniques for these steps were therefore improved upon and made more efficient as experiments progressed.

In all cases, the trunk of the human specimens without the heads, were snap frozen in OCT without fixation and kept at -80°C until sectioning. 7µm sections were then cryosectioned onto 5 PEN membrane slides for the initial CS17 experiments and 3 PEN slides for the later CS16 experiments with 6 sections on each slide. Each spatially defined sub-dissected region of interest is referred to as a sub-domain. Equivalent microdissected sub-domains from each tissue were pooled into the same collection tube. 6 slides were required initially to get a sufficient quantity of cells for earlier library-prep methods (detailed below) whereas later more efficient methods meant only 3 slides were required. Slide number was kept to a minimum to reduce microdissection time and prevent too much RNA degradation. Zeiss recommends 30 minutes maximum tissue exposure on the microscope stage which is roughly how long it takes to microdissect 3 slides. Additional slides increase quantity of RNA at the cost of RNA integrity.

The sections were stained with a rapid H&E stain and microdissected according to the dissection strategy for that experiment (detailed in section 3.2). Each sub-domain was dissected using LCM and collected into different collection tubes. The equivalent sub-domains from sister sections were pooled into the same tube and, when all slides had been completed, lysis buffer was immediately added. In order to get an idea of the RNA quality, large rectangular regions of tissue section (~150,000µm^2) were

also microdissected into another collection tube (Fig.10ai-ii). RLT lysis buffer from the RNeasy Micro RNA extraction kit was then added to extract RNA. This provides enough RNA for analysis on the TapeStation to get a RIN^e score (Fig.10b). The material collected in this tube had been exposed on the stage for longer than the microdissected subdomains so should represent the most degraded material. The RIN^e was also compared to the RNA RIN^e from a whole tissue section taken at the time of cryosectioning to assess how much degradation was due to the LCM procedure or due to degradation already in the initial frozen specimen. ΔRIN^e ranged from 0.2 to 1.9 indicating that the LCM protocol had minimal impacts on the RNA integrity of the samples.

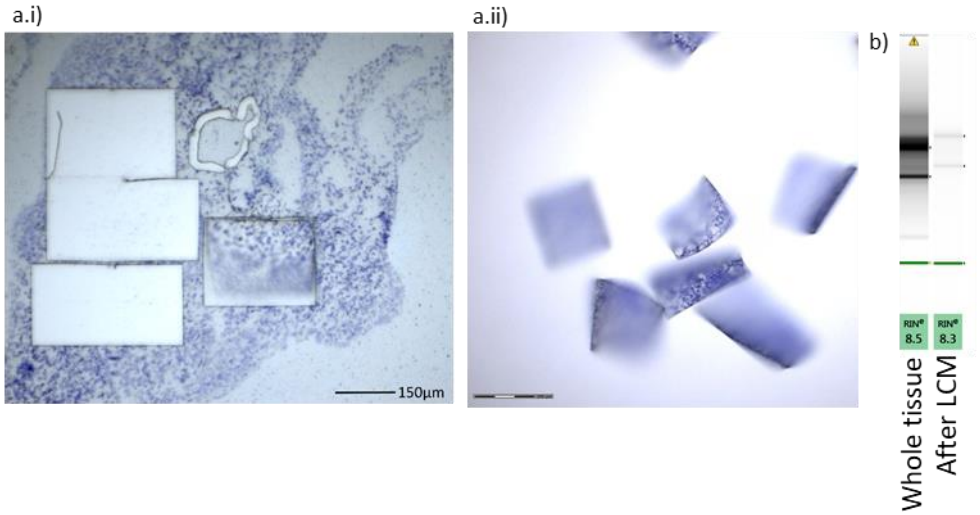


Figure 10: (a.i) Large rectangles of human LCM dissected material for measuring RIN^e of the tissue at the end of the LCM experiment, (a.ii) view of collected rectangles of tissue in the cap of the collection tube. (b) RIN^e of a whole tissue section taken during the cryosectioning step and after the complete LCM protocol prior to RNA-Seq library preparation**.** Electropherograms show the ratio between the 18S and 28S ribosomal RNA (rRNA) products**.**

For the CS17 experiments RNA was extracted and purified using the PicoPure RNA Isolation kit and the SMARTer® Stranded Total RNA-Seq Kit - Pico Input Mammalian Kit from Takara (version 1 for embryos n1-2, version 2 for sample n3) for RNA-Seq library preparation (Fig.11A). The SMARTer® kit manual suggests input of at least 250pg of RNA. For the initial experiment subdomains were pooled from 3 slides of 6 sections. ~70% of tissue sections were useable i.e. good morphology of cells and intact tissue architecture around the dorsal aorta therefore ~13 sections * ~25 cells per subdomain = ~325 cells per subdomain. 325 cells should give >3250pg of RNA.

The initial experiment included purified RNA from 2 LCM samples, a negative no RNA control and a 250pg of positive RNA provided in the kit. Library size distribution and quantification was assessed using the TapeStation (Fig.12a-d). A good library should be about 4000pg/ul and have a smooth curved

distribution peaking around 300 – 400bp. The negative control (Fig.12a) showed a small peak at 141bp indicating amplification of adapter dimers. The 2 LCM samples (Fig.12b, c) had extremely low quantities of 20.9 and 27.8 pg/ul whereas the positive control (Fig.12d) had 4330pg/ul with a smooth curve peaking at 311bp. This indicates that cells from microdissected tissue sections either do not yield as much RNA or that substantial RNA is lost during the RNA extraction and purification process. Increasing the slide number to 5 (~525 cells per subdomain) provided enough RNA to generate libraries (Fig.12e) although primer dimer peaks at around 140bp were often seen. Adapter dimers can be an indication of insufficient starting material but also that concentrations of adapter are too high and/or cDNA fragment size selection using Solid Phase Reversible Immobilisation (SPRI) beads insufficiently filtered out fragments of shorter length. Adapter dimers are sequenced so reduce the overall depth of the rest of the library. It is desirable therefore to remove them. Furthermore, after sequencing, these libraries showed very high levels of duplicates, often above 90% (Fig.14A). Duplicates are identical reads, which are unlikely to happen by chance. High levels can be caused by low complexity of the input RNA (low quantity of molecules) and PCR over amplification (Parekh *et al.*, 2016). Biases like these reduce the precision of differential gene expression (DGE) analysis. However, removal of duplicated reads does not improve accuracy of DGE and have been left in to the analysis (See Chapter 3 Discussion).

For the subsequent experiments with CS16 samples a different library preparation method was used taken from Nichterwitz et al, 2016 based on Smart-seq2(Picelli *et al.*, 2014) (Fig.11B). This uses individually purchased reagents rather than a kit which is more cost-effective. The mRNA is reverse transcribed and amplified directly from the hypotonic lysis buffer tested on mouse LCM samples (see section 3.1.2) which removes the requirement to extract and purify the RNA first which can lead to loss of RNA. Tagmentation (Tn5 transposase fragmentation), final amplification and barcoding is then done using the Illumina Nextera XT DNA sample preparation kit. To test the sensitivity of the protocol, libraries were prepared from regions of E11.5 mouse sections laser capture microdissected at increasing size; 20 cells (n=2), 10,000µm^2 (~66 cells, n=2), and 50000µm^2 (~330 cells, n=2) (Fig.13a). The pre-amplification step prior to fragmentation allows assessment of the success of protocol to that point and quality of the initial sample using the tapestation bioanalyzer. Mean library size should be around 1.5-2kb and lower base pair (bp) peaks indicates shorter fragments suggesting degradation of the initial sample (Picelli *et al.*, 2014). All samples had amplification products except for 50,000um^2 replicate (n) 1 (Fig.13ai-vi). However, the final tagmented/amplified and barcoded samples were also analysed on the TapeStation and showed successful libraries from all samples (Fig.13b) indicating that the absence of 50,000µm^2 n1 was a TapeStation issue. The pre-amplification curves had a size distribution peaking around 1.8kb showing that libraries could be generated from just 20 cells – a significant improvement on the SMARTer-Seq® method. Most curves were relatively free of shorter

fragments except for 50,000µm^2 n2 where a few shorter fragment peaks indicate some minor initial degradation of the RNA.

The final amplified libraries had smooth curves peaking at ~450bp although a primer dimer peak at around 40bp was also seen (Fig.13b). This peak was removed in subsequent experiments by more efficient SPRI bead size selection (Fig.13c) achieved by decreasing the ratio of beads to cDNA (0.8:1 to 0.7:1) which selects larger fragments. The quantities of the libraries were 0.8 – 0.96ng/µl (measured on a Qubit) showing good reproducibility of the tagmentation and final amplification steps. The output quantities were levelled as 1ng of pre-amplified cDNA is required as input to the tagmentation reaction for maximum efficiency.

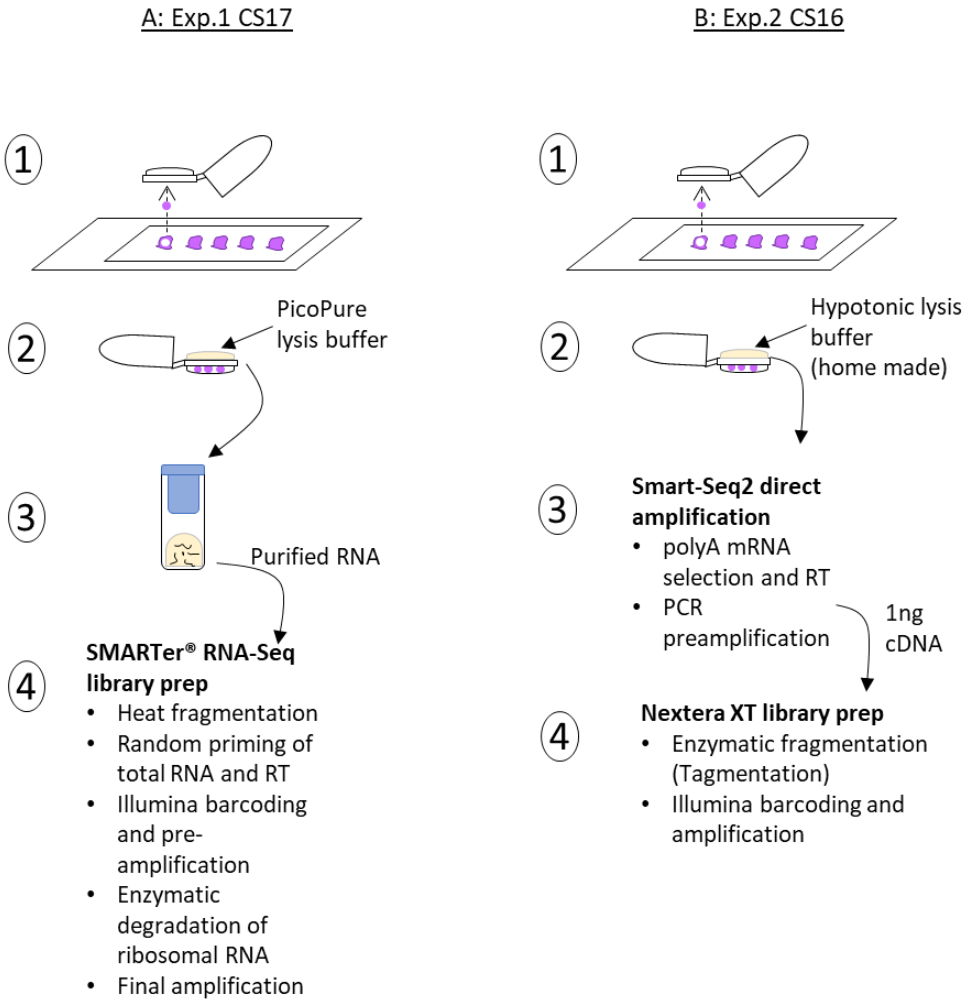


Figure 11 Schematic of protocol for two methods of RNA-Seq library generation from LCM material. A: 1) LCM material is collected in the cap of a collection tube. 2) Lysis buffer from the PicoPure kit is added on top of the material in the collection tube cap for cell lysis. 3) RNA is purified using the PicoPure RNA Isolation Kit. 4) RNA-Seq libraries are prepared using the SMARTer® Stranded Total RNA-Seq Kit – the main steps are provided in order. B: 1) LCM material is collected in the cap of a collection tube. 2) Hypotonic lysis buffer is added on top of the material in the collection tube cap for cell lysis. 3) The mRNA in the lysis buffer is directly reverse transcribed using oligo(d)T primers to select only polyA-tailed mRNA and pre-amplified. 4) The Nextera XT library prep kit is used to complete library preparation from the pre-amplified cDNA – the main steps are provided in order.

The protocol was shown to be effective down to at least 20 cells but to make sampling more equivalent across embryos, subdomains from 3 slides of 6 sections each were pooled to get signals averaged across the region. Reducing slide number to 3 slides helped streamline the LCM process and prevent unnecessary RNA degradation. This provided extra time to take additional subdomains from the tissue sections. The libraries generated had smooth peaks with average ~350 – 400bp and no adapter dimer peaks (Fig.13c). After sequencing, a lower percentage of reads were detected as duplicates than from the SMARTer-Seq® kit prepared libraries (Fig.14B). However, as fragmentation in the second protocol occurs after pre-amplification, duplicated reads may be fragmented in different positions and therefore not detected as duplicates computationally.

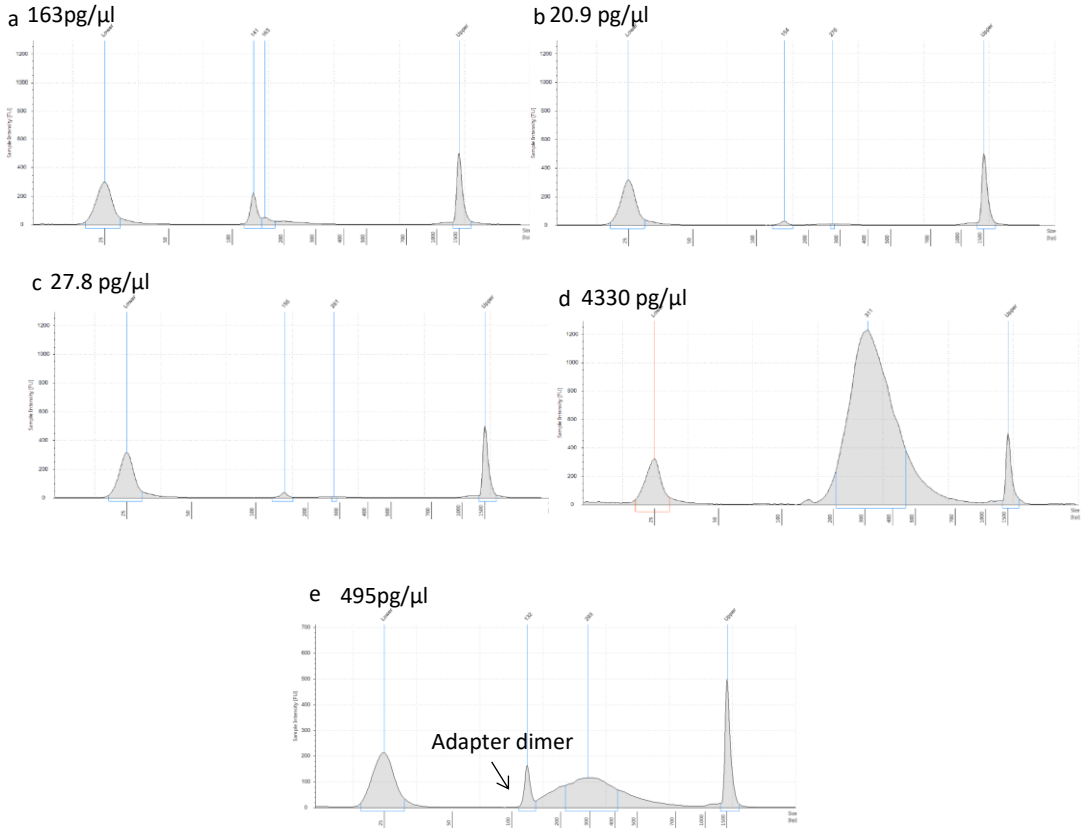


Figure 12. SMARTer® RNA-Seq library QC and quantification (a-d) shows library size distributions after preliminary CS17 LCM-Seq library preparation using the SMARTer® Stranded Total RNA-Seq Kit - Pico Input Mammalian Kit; (a) is a negative control, (b,c) are ~325 cells from CS17 LCM microdissected dorsal aortas. (d) is a positive control of 250pg initial RNA quantity. (e) representative image of the library size distribution prepared from ~525 cells of CS17 LCM microdissected dorsal aorta

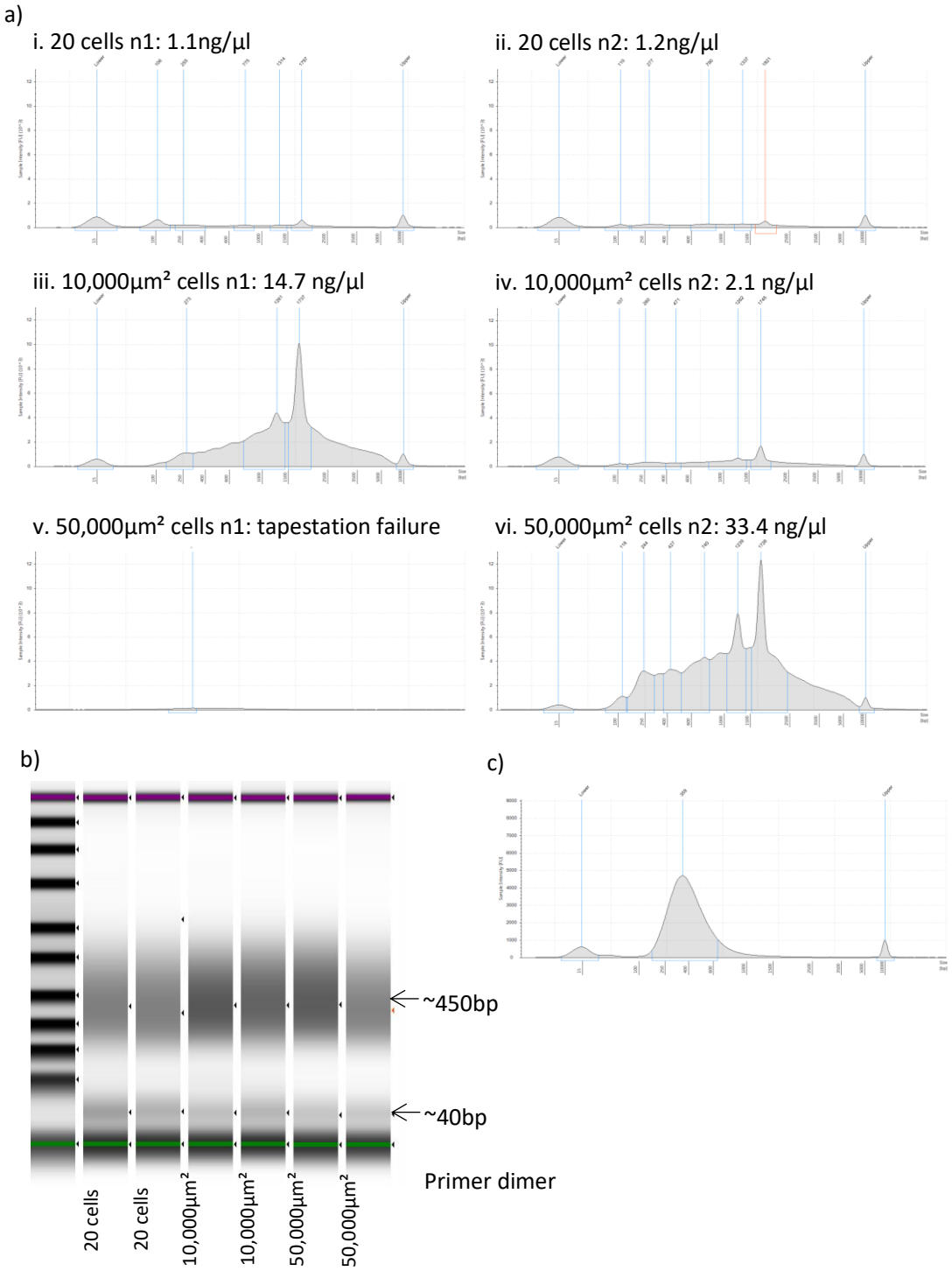


Figure 13. Smart-Seq2 RNA-Seq library QC and quantification. (a) Fragment size and distribution following first-round amplification of E11.5 LCM cells using the Smart-Seq2 protocol. n = replicate number. Peaks marked as lower and upper are tapestation markers and not cDNA. (b) Library size distributions following tagmentation and final amplification steps – consistent peak at ~450bp for all samples and ~40bp primer dimer peak. (c) Representative library size distribution plot showing no primer dimer peak after more efficient SPRI bead selection

A.

Sample Name	%Dups	Cycles
N1 Ventral R1	94.30%	33
N1 Ventral R2	76.30%	33
N1 Ventro-lateral R1	95.00%	33
N1 Ventro-lateral R2	75.70%	33
N1 Dorsal-lateral R1	94.10%	33
N1 Dorsal-lateral R2	77.40%	33
N1 Dorsal R1	79.80%	21
N1 Dorsal R2	67.60%	21
N2 Ventral R1	89.00%	23
N2 Ventral R2	77.60%	23
N2 Ventro-lateral R1	85.50%	23
N2 Ventro-lateral R2	77.50%	23
N2 Dorsal-lateral R1	94.90%	32
N2 Dorsal-lateral R2	80.80%	32
N2 Dorsal R1	83.90%	27
N2 Dorsal R2	74.20%	27
N3 Ventral L1 R1	91.10%	27
N3 Ventral L3 R2	67.10%	27
N3 Ventro-lateral L1 R1	93.90%	27
N3 Ventro-lateral L1 R2	74.70%	27
N3 Dorsal-lateral L1 R1	92.50%	27
N3 Dorsal-lateral L1 R2	69.90%	27
N3 Dorsal L1 R1	89.20%	27
N3 Dorsal L1 R2	68.90%	27
N3 Ventral L2 R1	91.50%	27
N3 Ventral L2 R2	65.90%	27
N3 Ventro-lateral L2 R1	94.30%	27
N3 Ventro-lateral L2 R2	73.80%	27
N3 Dorsal-lateral L2 R1	92.90%	27
N3 Dorsal-lateral L2 R2	69.00%	27
N3 Dorsal L2 R1	89.70%	27
N3 Dorsal L2 R2	67.30%	27

B.

Sample Name	% Dups	Cycles
N1 Inner Dorsal R1	63.80%	26
N1 Inner Dorsal R2	59.30%	26
N1 Inner Ventral R1	61.50%	24
N1 Inner Ventral R2	56.00%	24
N1 Mid Dorsal R1	58.80%	26
N1 Mid Dorsal R2	55.20%	26
N1 Mid Ventral R1	56.70%	24
N1 Mid Ventral R2	50.80%	24
N1 Outer Dorsal R1	54.10%	26
N1 Outer Dorsal R2	51.40%	26
N1 Outer Ventral R1	58.10%	26
N1 Outer Ventral R2	54.70%	26
N2 Inner Dorsal R1	67.60%	29
N2 Inner Dorsal R2	62.40%	29
N2 Inner Ventral R1	75.90%	29
N2 Inner Ventral R2	67.80%	29
N2 Mid Dorsal R1	53.20%	29
N2 Mid Dorsal R2	49.50%	29
N2 Mid Ventral R1	51.70%	25
N2 Mid Ventral R2	48.50%	25
N2 Outer Dorsal R1	29.30%	25
N2 Outer Dorsal R2	26.00%	25
N2 Outer Ventral R1	38.00%	25
N2 Outer Ventral R2	34.40%	25
N3 C1 Inner Ventral R1	59.50%	24
N3 C1 Inner Ventral R2	58.20%	24
N3 C1 Inner Dorsal R1	61.10%	24
N3 C1 Inner Dorsal R2	60.00%	24
N3 C1 Mid Ventral R1	57.90%	23
N3 C1 Mid Ventral R2	56.30%	23
N3 C1 Mid Dorsal R1	55.90%	24
N3 C1 Mid Dorsal R2	55.10%	24
N3 C1 Outer Ventral R1	60.50%	23
N3 C1 Outer Ventral R2	60.00%	23
N3 C1 Outer Dorsal R1	58.70%	24
N3 C1 Outer Dorsal R2	58.00%	24
N3 C2 Inner Ventral R1	57.00%	24
N3 C2 Inner Ventral R2	56.30%	24
N3 C2 Inner Dorsal R1	58.00%	24
N3 C2 Inner Dorsal R2	56.60%	24
N3 C2 Mid Ventral R1	58.20%	24
N3 C2 Mid Ventral R2	57.50%	24
N3 C2 Mid Dorsal R1	60.20%	24
N3 C2 Mid Dorsal R2	59.20%	24
N3 C2 Outer Ventral R1	54.90%	24
N3 C2 Outer Ventral R2	54.30%	24
N3 C2 Outer Dorsal R1	57.90%	23
N3 C2 Outer Dorsal R2	57.50%	23

Figure 14. Percentage of sequenced reads that are duplicates and number of amplification cycles for each sample in CS17 (A) and CS16 (B) LCM-Seq experiments. R1 = Read 1, R2 = Read 2. Sample number (N) is also indicated.

3.2 Spatial transcriptome profiling of the CS17 dorsal aorta

3.2.1 CS17 LCM-microdissection strategy

To investigate dorsal-ventral (D-V) differences in gene expression and signalling patterns in the dorsal aorta, 3 CS17 human embryos were transcriptionally profiled. In order to take an equivalent region from each embryo along the rostro-caudal axis and thus reduce variation, anatomical landmarks in the histology sections were used as a point of reference. Sections for LCM were taken in the pre-liver region between the midgut loop (caudal limit) and the duodenum/superior mesenteric artery (Fig.15a-bii) as this region is reported to have a high number of IAHCs during the HSC developmental window (Tavian *et al.*, 1996; Tavian, Hallais and Peault, 1999).

The LCM microdissection strategy was to sub-dissect the wall of the dorsal aorta, composing the endothelial and perivascular layer, into 4 sub-domains along the D-V axis including ventral (V), ventro-lateral (VL, 1 from each side), dorsal-lateral (DL, 1 from each side) and dorsal (D) (Fig.15c). As the endothelial cells are not immune-labelled it is likely that additional mesenchymal cells are included in the sub-dissected regions. A pink layer of eosin staining around the dorsal aorta was used to guide dissections of the vessel wall (Fig.15d.ii). Sub-dissecting in the exact same pattern between embryos and even tissue sections of the same embryo was made complicated due to differing morphologies of the dorsal aorta (Fig.15di-iv) and branching vessels. Although the sub-domains vary slightly in position from section to section and their boundaries may overlap, pooling same sub-domains across tissue sections will average the signals and is expected to strengthen polarised differences.

3.2.2 Validation of the presence of IAHCs in the ventral domain of the dorsal aortas within the regions taken for LCM-Seq

The goal of finding D-V polarised gene expression is to identify signals that may promote haematopoiesis in the AGM region. It was therefore necessary to validate the presence of IAHCs in the ventral domain of the dorsal aorta in the regions subdissected for RNA sequencing. During cryosectioning for LCM, when one PEN membrane slide had been completed an additional section was transferred to a glass slide for immunohistochemistry. These sections were stained with antibodies for CD31 and Runx1 to label IAHCs. In each embryo there is a ventrally localised IAHC (Fig.16). Embryos N2 and N3 have quite large IAHCs whereas in N1 only 2 cells are visible. This might indicate that N1 is a slightly later stage than N2/3 or simply that a larger IAHC hasn't been captured in this section. It could also be due to genetic variation between samples. As ventral IAHCs were confirmed in all 3 embryos within the regions taken for LCM all samples were submitted for sequencing.

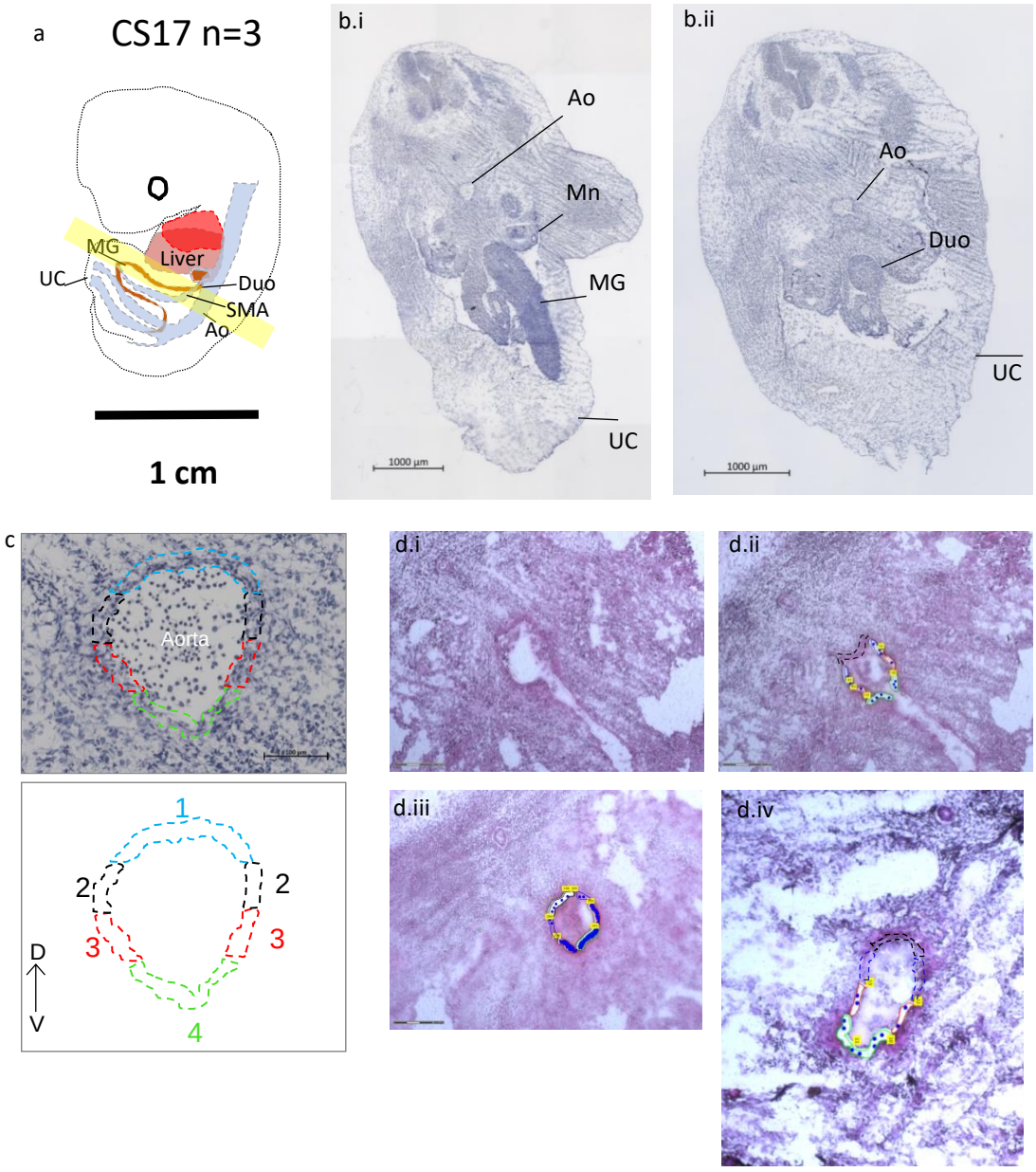


Figure 15. CS17 LCM-Seq microdissection strategy. (a) CS17 embryo schematic with yellow bar indicating the region within which sections were taken for LCM-Seq. Ao = Dorsal aorta, Duo = duodenum, UC = umbilical cord, MG = mid gut, SMA = superior mesenteric artery. (b) The histology images indicate the caudal (b.i) and rostral (b.ii) limits of the embryo taken for LCM-Seq with the use of anatomical landmarks to define position along the rostral-caudal axis. Ao = Dorsal Aorta, Mn = Mesonephros, MG = Mid Gut, UC = Umbilical cord, Duo = Duodenum. (c) shows the CS17 LCM-Seq microdissection strategy of 4 domains along the dorso-ventral axis (1 = Dorsal, 2 = Dorsal-lateral, 3=Ventral-lateral, 4=Ventral). (d) Representative images from the PALM laser capture microscope of the CS17 dorsal aorta from different sections. Any coloured lines mark the sub-domains directing the laser where to cut.

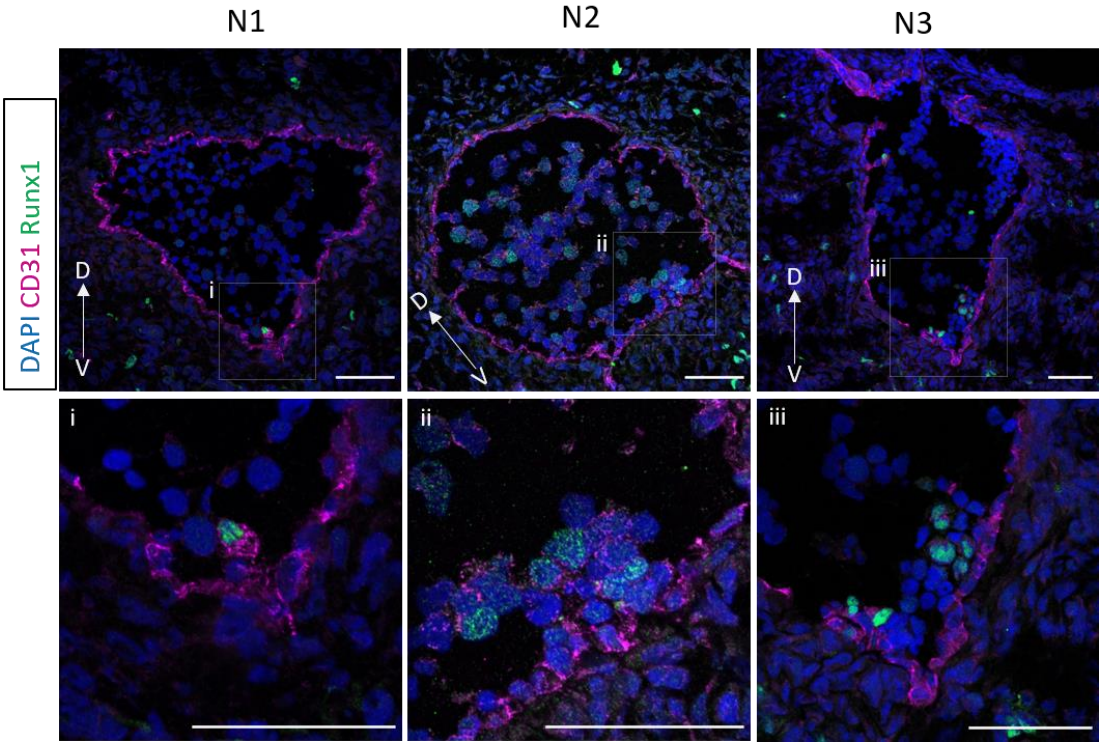


Figure 16 Validation of the presence of IAHCs in the ventral domain of the CS17 dorsal aortae. Dorsal aortae from sister sections of sections taken for the CS17 LCM-Seq experiment. Ventrally localised clusters are validated by their expression of Runx1. (i-iii) are zoomed in images of regions indicated by squares in the top row. Nx = replicate number. Scale bars are 50µm. D = Dorsal, V = Ventral.

3.2.3 Visualisation and quality assessment of the transcriptome by sample clustering analysis

The principle component analysis of the RNA-Seq datasets shows no clear clustering of samples by subdomain or embryo number (Fig.17ai-ii). It is of note that samples V2, D1, D2 and VL2 were shifted to the positive loading of PC1 and these samples had a mean of 8.5% less duplicates than the other samples (Fig.14A) indicating that lower levels of duplicates may allow other genes to be more highly expressed. This pattern of clustering by duplicate level is seen on the heatmap of top 500 genes by expression level (Fig.17b). The heatmap clusters into 3 main groups which corresponds to duplicate level ranges of 79.8 – 89% (V2, D1, D2) 89.7 – 94.1% (V3, DL3, D3, DL1) and 94.3 – 95% (V1, VL1, VL3, DL2). Furthermore, the heatmap of sample-to-sample distances shows clustering of the 2 groups with lower duplicate ranges, whereas the group with high duplicate levels have low similarity to the other samples indicating high duplicate level may obscure true variations in gene expression (Fig.17c). VL2 clusters alone, which may be to do with the much higher sequencing depth for this sample compared to others (data not shown). This high sequencing depth probably resulting from an erroneously high input amount of this sample for sequencing. The lack of expected clustering proved not critical as shown in further analysis (Section 3.2.4 – 3.2.7).

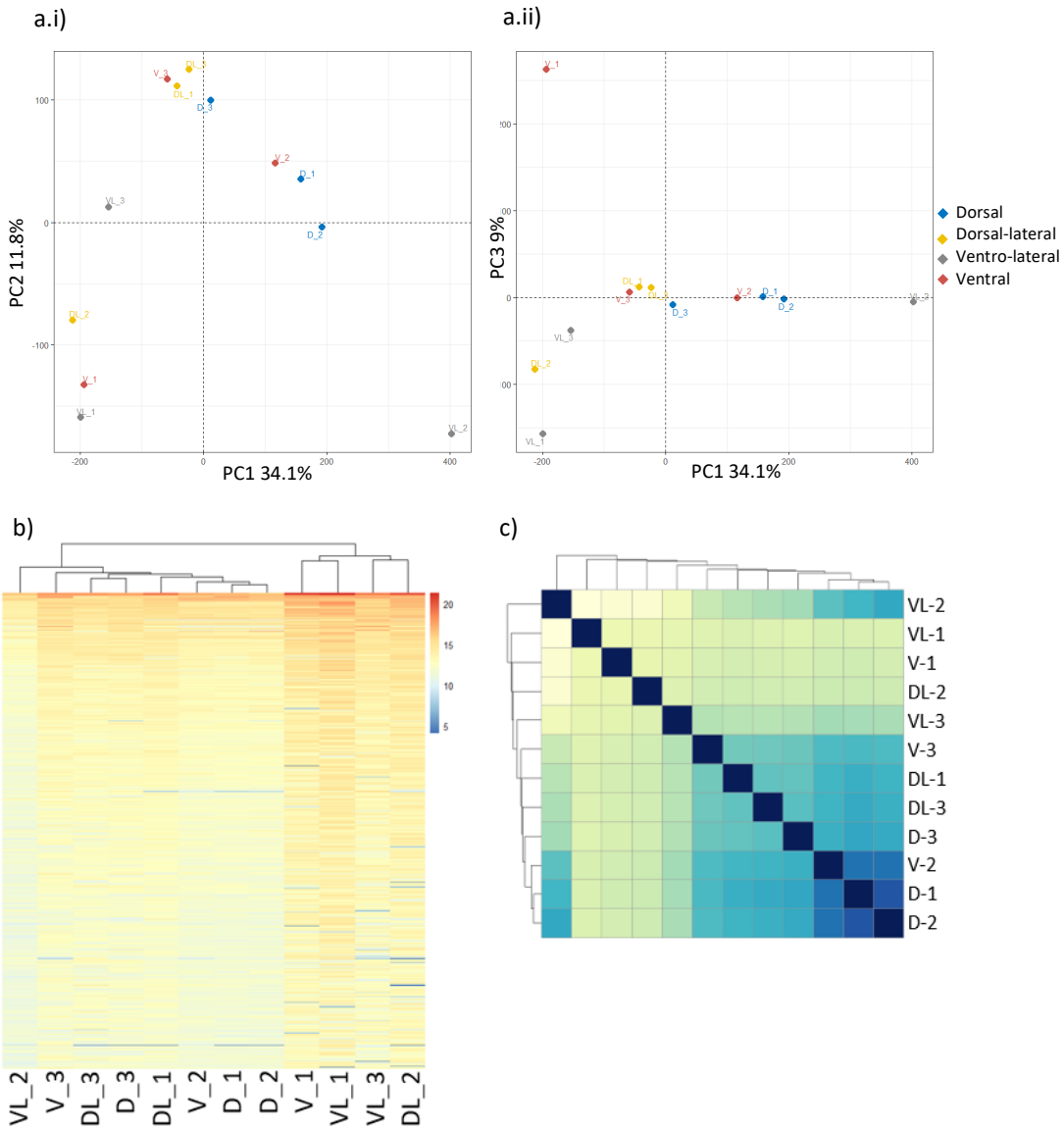


Figure 17. CS17 LCM-Seq visualisation and clustering of samples a) Principle component analysis (PCA) of CS17 LCM-Seq showing PC1 and PC2 (a.i) and PC1 and PC3 (a.ii). (b) Heatmap of the 500 top genes by mean expression level with hierarchical sample clustering. (c) Heatmap of sample-to-sample distances with hierarchical clustering.

3.2.4 Differential gene expression analysis reveals genes enriched for each subdomain along the dorsal-ventral axis

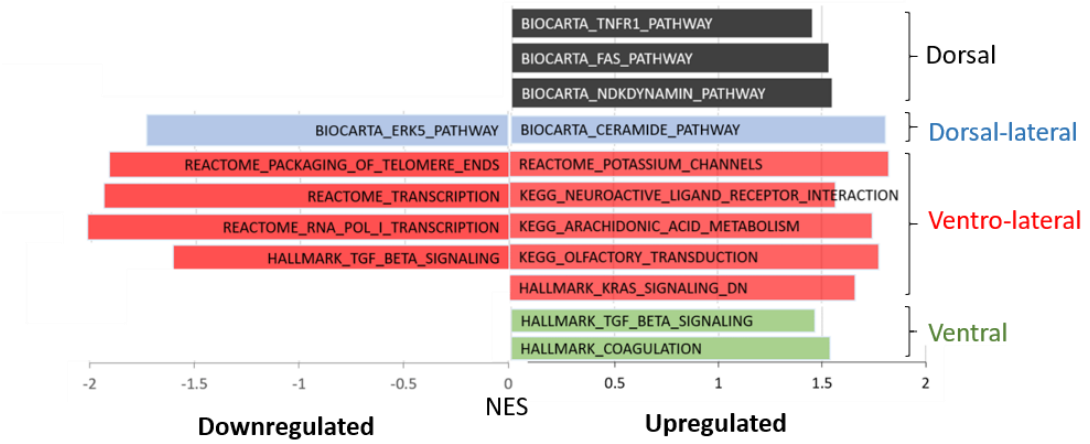
Differential gene expression analysis was carried out for each subdomain vs. the other 3 subdomains and revealed significant positive upregulation for genes in each subdomain. D had a particularly high number of enriched genes at 774, with 352, 64, and 159 enriched genes for the DL, VL and V domains respectively (Fig. 18b). To identify pathways enriched for each sub-domain, Gene set enrichment analysis (GSEA) was carried out on differentially expressed (DE) genes ranked by log2 fold change using Hallmark, BioCarta, Reactome and KEGG databases (Fig.18a). 'TNFR1' and 'FAS' pathways enriched in D, and 'Ceramide' pathway enriched in DL are all involved in regulation of apoptosis, specifically induction of apoptosis. 'TGFβ signalling', which is known to be involved in promoting HSC emergence, is enriched in V but down-regulated in VL suggesting a more confined localisation of this signalling to the most ventral part of the dorsal aorta. VL is also enriched for 'neuroactive ligand receptor interactions', which is consistent with the VL localisation of sympathetic ganglia known to be an important component of the AGM HSC niche (Fitch *et al.*, 2012). This analysis highlights that the dorsal aorta is a dynamic niche with different signalling roles along the D-V axis.

3.2.5 Gradients of gene expression along the dorsal-ventral axis

To identify gradated changes in gene expression along the D-V axis using RNA-Seq analysis of the sub-dissected regions, a likelihood ratio test (LRT) was used instead of the default Wald test during DESeq2 (Love, Huber and Anders, 2014). This tool is frequently used for elucidating changes in gene expression along a time series, so here, by modelling the axis of the dorsal aorta as a time series from dorsal to ventral or vice versa, DE genes along the D-V axis could be revealed.

This analysis gave only 3 significant genes with expression gradients along the D-V axis ($p.adj<0.05$, Fig.18c). In the dorsal-to-ventral direction there was the highly significant gene *REN*, which encodes the blood pressure regulator Renin (REN) and the gene *SLC26A7*, which encodes the sulfate/anion transporter Solute Carrier Family 26 Member 7. In the ventral-to-dorsal direction, the gene *PCDH18* encoding Protocadherin 18 increases in expression.

a)



b)

Subdomain	No. genes positively enriched
Dorsal (D)	774
Dorsal-lateral (DL)	352
Ventro-lateral (VL)	64
Ventral (V)	159

c)

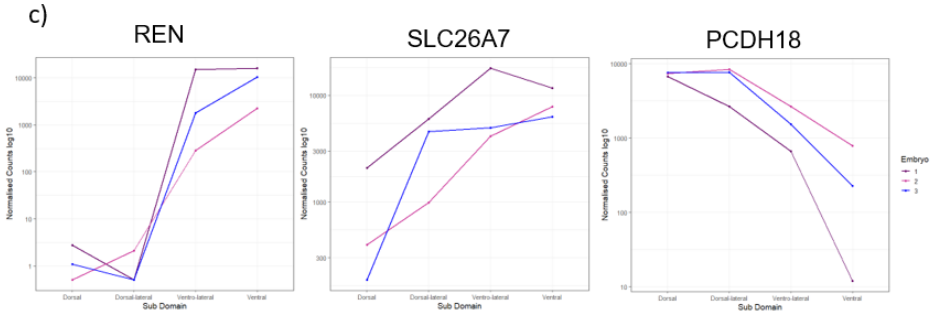


Figure 18 : Differentially expressed genes and pathways along the CS17 dorsal aorta dorsal-ventral axis. a) GSEA of differentially expressed genes for each sub-domain vs. the other 3. X-axis shows normalised enrichment score (NES) with positive values for upregulated pathways and negative values for downregulated pathways (b) Number of genes significantly upregulated in each domain vs. the other 3, p.adj<0.05. (c) Genes significantly differentially expressed along the dorsoventral axis (test = Likelihood Ratio Test (LRT), p.adj<0.05)

3.2.6 Pooled ventral and dorsal analysis reveals spatially polarised gene expression patterns

To determine whether a simpler D-V split would reveal additional spatially polarised genes and pathways, D and DL and V and VL subdomains were combined to dorsal_all and ventral_all domains respectively (Fig.19a). 93 genes were significantly upregulated in the ventral_all domain compared to 825 genes in the dorsal_all domain. The top ventral gene was the lncRNA *ENSG00000279310*. *REN*, which was previously revealed as a gene with a gradient of gene expression towards the ventral domain (section 3.2.5), was the next most enriched gene in the ventral_all domain (Fig.19c,d).

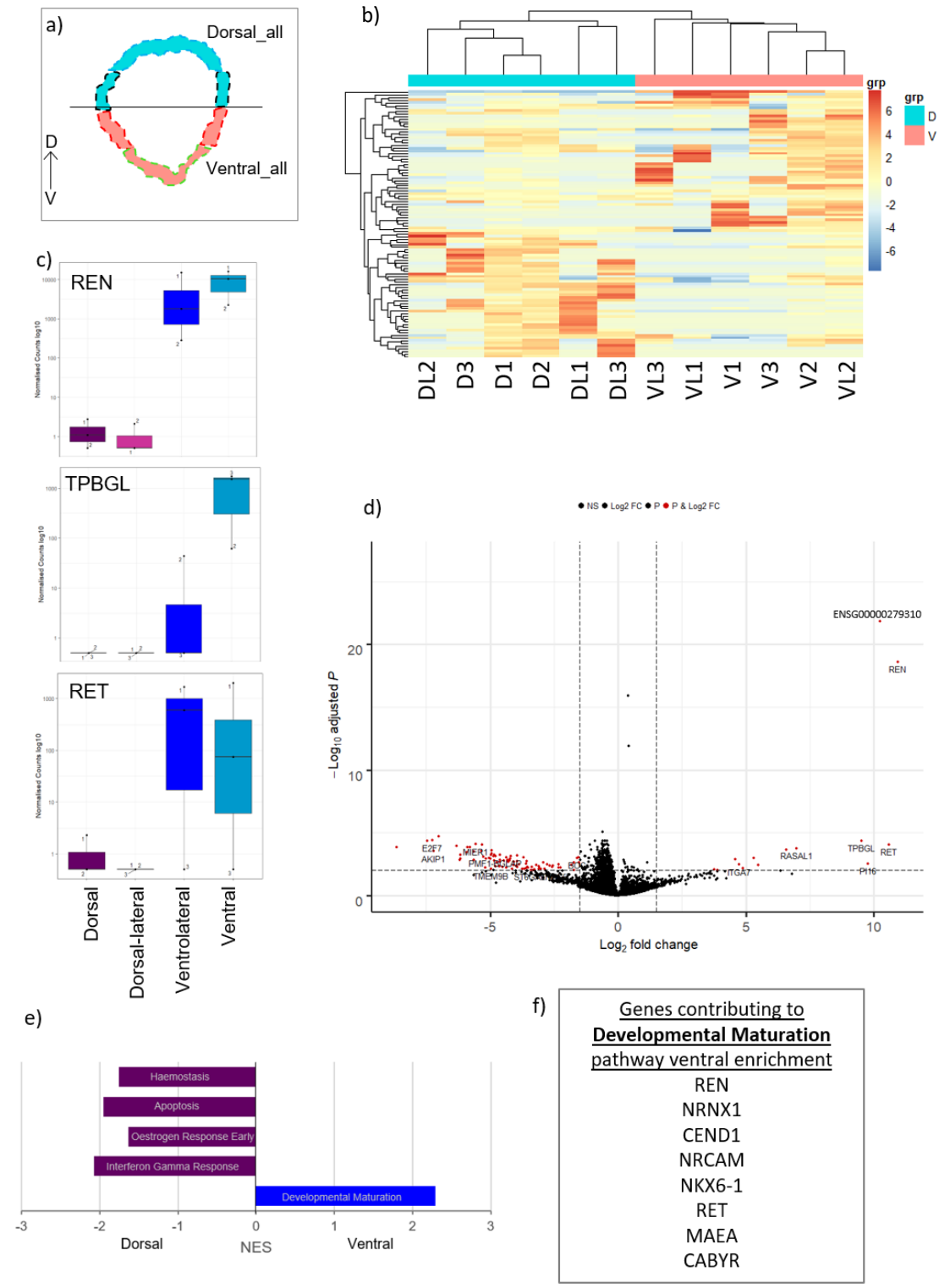


Figure 19: CS17 ventral_all vs dorsal_all analysis. a) Grouping of ventral and ventro-lateral subdomains (ventral_all) and dorsal and dorso-lateral (dorsal_all) for differential gene expression analysis. (b) Heatmap of top 200 differentially expressed (DE) genes between ventral_all and dorsal_all. (c) 3 top ventral_all upregulated genes vs dorsal_all (p.adj<0.05). (d) Volcano plot of DE genes between ventral_all and dorsal_all. In red = Log_2 fold change > 1.5 and p.adj<0.05 (e) GSEA of DE genes between ventral_all and dorsal_all. f) Genes contributing to ventrally enriched Developmental Maturation pathway, GO: 0021700

GSEA showed four pathways enriched in the dorsal_all domain compared with just one in the ventral_all domain (Fig.19e). 'Apoptosis' pathway enrichment in the dorsal_all domain is not surprising since this pathway was enriched in both D and DL sub-domains separately. The other three pathways enriched in dorsal_all are 'Haemostasis', 'Oestrogen Response', and 'Interferon Gama Response'. The only pathway enriched in the ventral_all domain was the Gene Ontology (GO) pathway 'Developmental Maturation' (GO:0021700). This pathway describes broadly "A developmental process, independent of morphogenetic (shape) change, that is required for an anatomical structure, cell or cellular component to attain its fully functional state" and includes top DE genes such as *REN* and *RET* (Fig.19f).

In total, fewer genes and pathways were found from the ventral_all vs. dorsal_all analysis compared with the four sub-domain analyses. This indicates the utility of precisely resolving the spatial transcriptome to smaller regions along the D-V axis as it reveals additional localised signalling pathways.

3.2.7 Correcting for duplicate clustering effect on differential gene expression between dorsal and ventral domains of the dorsal aorta

The DGE analyses in sections 3.2.4 – 3.2.6 were carried out by first correcting for per embryo differences in gene expression before contrasting subdomains. As the sample clustering analysis described in section 3.2.3 indicated potential batch bias by levels of duplication a separate analysis was run correcting first for duplication level batch effects instead of by embryo to see if it provided more meaningful data.

DGE analysis of ventral_all vs. dorsal_all gave 181 and 147 upregulated genes respectively ($p.adj<0.05$) indicating that the high number of dorsal_all genes in section **3.2.6** may have been influenced by duplicate bias. The most significantly (i.e. smallest adjusted p.value) upregulated for ventral_all was PI16, Peptidase Inhibitor 16 which was also significantly ventrally upregulated in the previous analysis (Fig.20a, b).

GSEA of the enriched genes showed no pathway enrichment. As correcting for batch based on duplicate level rather than embryo didn't give much new meaningful data, the 'per embryo' batch corrected dataset is considered the principal analysis. This may risk false positives, particularly with the high number of genes upregulated in the dorsal_all domain but this was preferable to losing data through over-conservative batch correction. Furthermore, the elucidation of pathways in the original analysis indicates concordant DE of genes within pathways whereas this analysis gave no pathway enrichment, weakening the validity of the dataset.

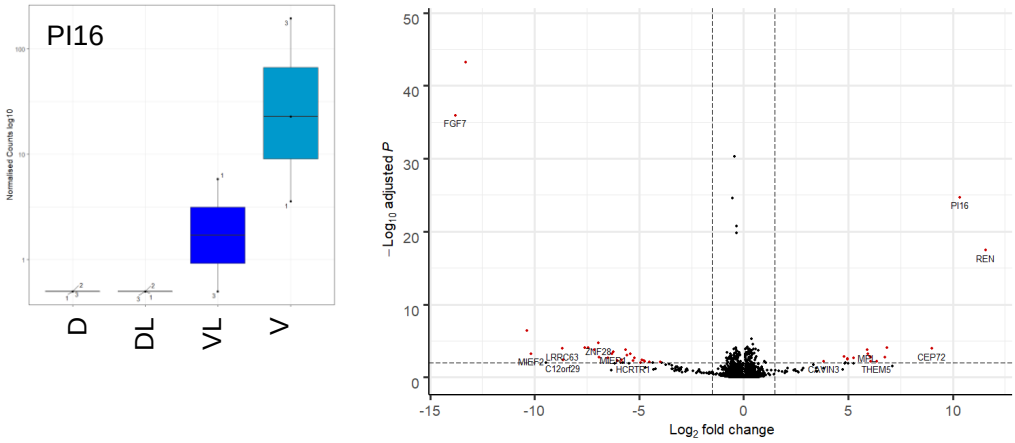


Figure 20: : CS17 ventral_all vs dorsal_all differential expression analysis following removal of duplicate bias effects (a) Most significant (p.adj = 2.29E-24) ventral gene, PI16 and (b) volcano plot of differentially expressed genes

3.3 Spatial transcriptome profiling of the CS16 AGM region

3.3.1 CS16 LCM microdissection strategy

Three CS16 embryos were used to spatially map the transcriptome of the human AGM region with LCM-Seq. Three concentric rings radiating away from the dorsal aorta were microdissected and for each ring the dorsal and ventral domains were sub-dissected and collected separately (Fig.21c). This approach was designed to allow signalling pathways for different cell layers of the HSC niche to be revealed, proximal and distal to the IAHCs, as well as to investigate D-V differences. The 3 layers are referred to as Inner (I), Mid (M) and Outer (O) which contain mainly I) endothelial and perivascular cells, M) sub-aortic mesenchymal cells and O) cells termed 'outer mesenchymal' cells respectively (Fig.21c, Table 7). Due to the nature of the technique these layers will not be made exclusively of these cell types. The Outer layer will likely include some endothelial cells of branching vessels and macrophages captured in the sub-aortic mesenchyme. This will be taken into account when analysed. However, in this project the aim is to determine primarily the proximity of the signalling to the dorsal aorta and the spatial polarisation of the signalling. Signals can be resolved to cell types through later experiments.

Dorsal/Ventral	Layer	Description	Shorthand
Ventral	Inner	Endothelial/Perivascular	V_Inner
Dorsal	Inner	Endothelial/Perivascular	D_Inner
Ventral	Mid	Sub-aortic mesenchyme	V_Mid
Dorsal	Mid	Sub-aortic mesenchyme	D_Mid
Ventral	Outer	Outer mesenchyme	V_Outer
Dorsal	Outer	Outer mesenchyme	D_Outer

Table 7: Sub-domains taken for CS16 LCM-Seq. Description of samples with the shorthand used throughout text.

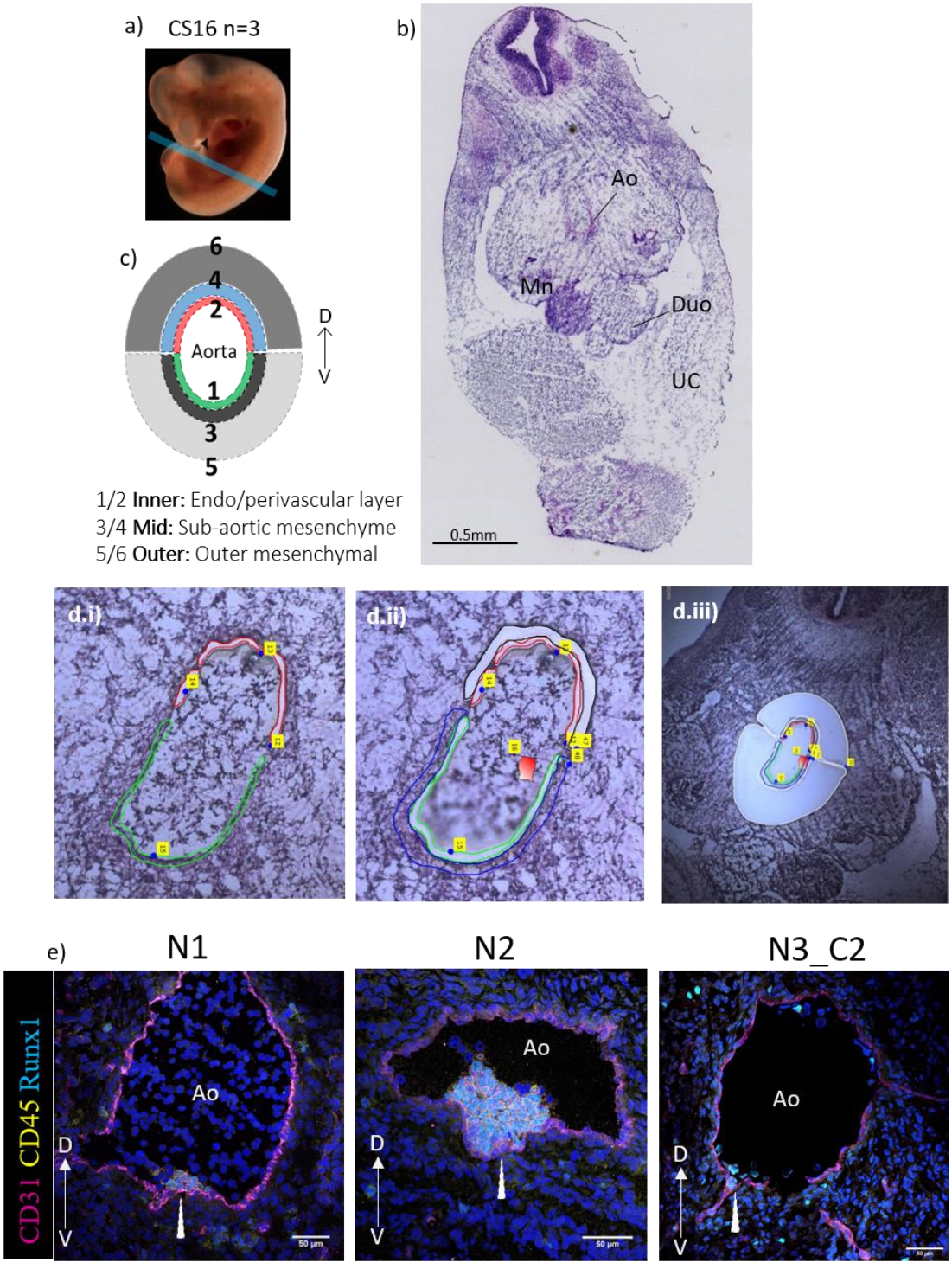


Figure 21 Sub-dissection strategy for LCM-Seq of the CS16 dorsal aorta and surrounding mesenchyme (a) Representative CS16 embryo. The blue line indicates the boundary of the region taken for LCM-Seq transcriptome analysis of the CS16 AGM. (b) The histology image indicates anatomical features used to define correct region for LCM along the rostral – caudal axis. (Ao = Dorsal Aorta, Mn = Mesonephros,, UC = Umbilical cord, Duo = Duodenum). (c) shows the CS16 LCM-Seq microdissection strategy (d) Representative images from the laser capture microscope of the CS16 AGM indicating Inner (i), Mid (ii), and Outer (iii) layers. (e) Immunostaining for CD31, CD45 and Runx1 validates the presence of ventralised intra-aortic haematopoietic clusters (IAHCs) in each of the CS16 dorsal aortae taken for LCM-Seq. White arrowhead indicates IAHC. Ao = Dorsal aorta. D=Dorsal V=Ventral. Scale bar = 50µm.

Embryo N3 had two regions along the rostral-caudal axis taken for LCM-Seq. As the embryos were supplied externally and arrived pre-embedded, exact orientation of the embryo in the OCT block was not controlled. This made it difficult to determine the exact anatomical location of the first sections taken for LCM-Seq (C1) and upon staining it appeared that they were too caudal in position and were lacking obvious Runx1+ IAHCs (Appendix Fig.1). More rostral sections were taken for the second LCM-Seq (C2), at the region of the vitelline artery, and this time ventral IAHCs were verified (Fig.21e, N3_C2). cDNA libraries were sequenced from both LCM-Seq libraries for this embryo and included in the principal component analysis. Only N3_C2 was used for differential expression analysis.

3.3.2 Validation of the presence of IAHCs in the ventral domain of the dorsal aortae within the regions taken for LCM-Seq

Once again, sister sections of those microdissected for LCM-Seq were stained for CD31, Runx1 and CD45 to verify the presence of IAHCs in the ventral domain. Each embryo showed ventrally localised IAHCs (Fig.21e) with the exception of embryo N3_C1, as discussed above, which was not included in the DGE analysis. As with CS17, the IAHCs varied in size between embryos which may be due to genetic variability or just variable sizes of IAHCs along the rostral-caudal axis. Although Runx1+ cells were seen in the dorsal domain (Fig.21e, N3_C2), these rarely co-expressed endothelial CD31 and never seemed integrated into the endothelium indicating that IAHCs are exclusively found in the ventral domain.

3.3.3 Visualisation and quality assessment of the transcriptome by sample clustering analysis

CS16 LCM-Seq sample similarity was analysed using dimension reduction principal component analysis (PCA). Plotting of the first 2 principal components (PCs) show clustering by embryo number indicating that biological variability is stronger than variability between subdomains (Fig.22ai). This is also visible on the sample distances heatmap with a large central cluster of N3 samples (both C1 and C2), a smaller left-hand cluster for N2 and the remaining samples belonging to N1 with the exception of VM-2 (Fig.22b).

PC2/PC3 plotting shows clustering of the samples into layers with Inner samples shifted towards the PC3 negative loading, Mid samples around 0 and Outer samples shifted towards the positive loading (Fig.22aii). This indicates that within the AGM niche there are robust differences in gene expression depending on proximity to the dorsal aorta.

Expression levels of endothelial and perivascular markers were visualised for all samples to validate the cell types within each layer (Fig.22c). The Inner layer had the highest expression of the endothelial markers CD31 (*PECAM1*), VE-Cadherin (*CDH5*) and *CD34* as expected as well as highest expression of perivascular cell markers smooth muscle alpha (α)-2 actin (*ACTA2*), cardiac muscle alpha (α)-1 actin (*ACTC1*), and NG2 (*CSPG4*). Mid layer samples had lower expression of endothelial markers than Inner

but still expressed perivascular markers indicating either that the Mid layer also contains perivascular cells or that the sub-aortic mesenchyme also expresses these markers. Outer layer samples had low expression of both endothelial and perivascular markers as expected.

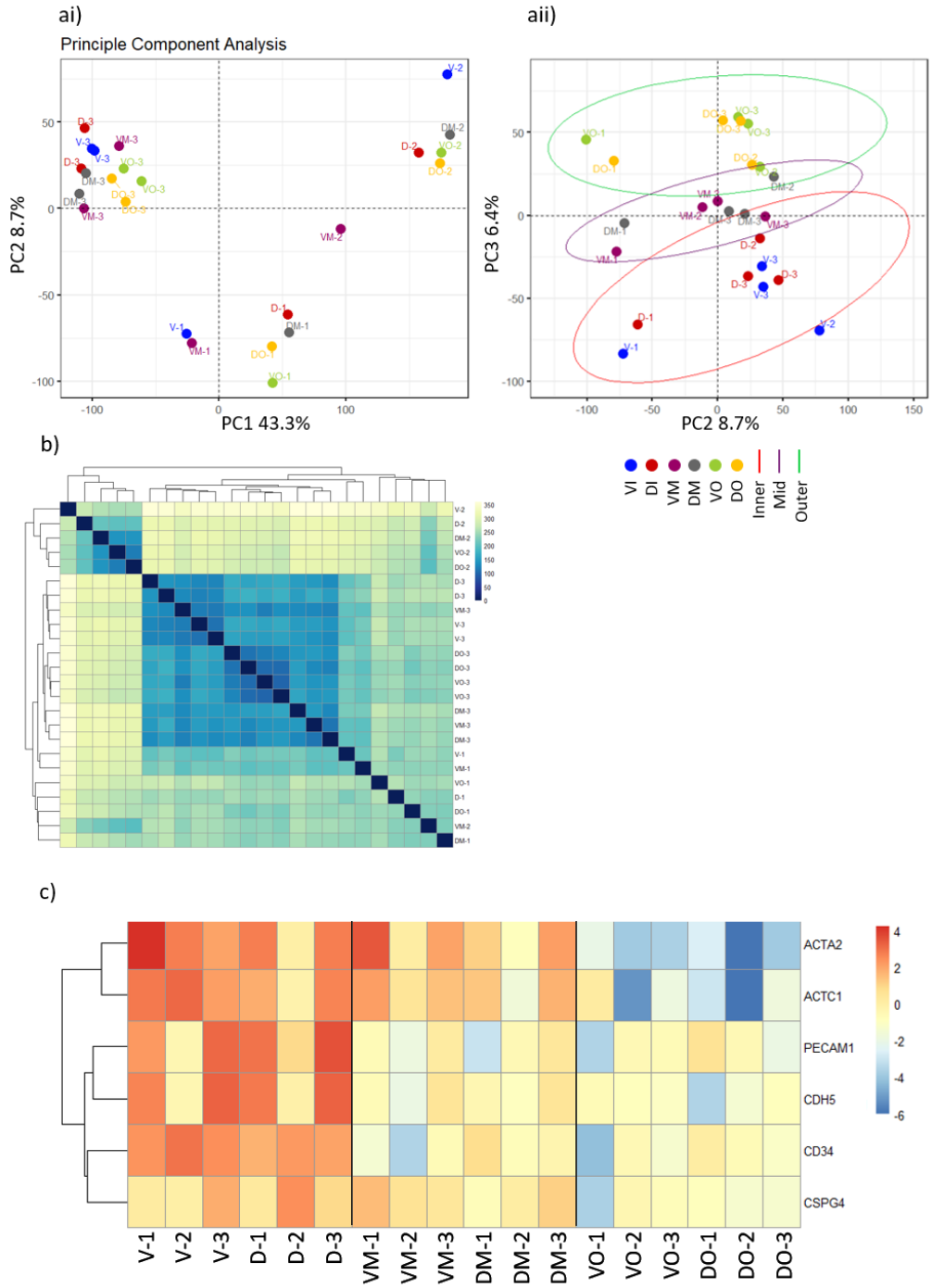


Figure 22. CS16 LCM-Seq visualisation and clustering of samples: (a) PCA; (ai) shows PC1 and PC2, (aii) shows PC2 and PC3. (b) heatmap of sample-to-sample distances with hierarchical clustering. (c) heatmap of normalised expression for select endothelial and perivascular cell markers.

3.3.4 Spatial dorsal-ventral gene expression polarisation across the dorsal aorta and surrounding mesenchymal tissues

DGE analysis was used to compare the dorsal and ventral subdomains for each subdissected layer (Inner, Mid, Outer), (Table 7). The numbers of significantly (p.adj<0.05) enriched genes for each concentric layer (subdomain) were:

- Inner: 52 Ventral and 53 Dorsal
- Mid: 129 Ventral and 100 Dorsal
- Outer: 17 Ventral and 47 Dorsal

There is only one gene with consistently polarised expression in every layer, the forkhead transcription factor Forkhead Box F2 (*FOXF2*). *FOXF2* is upregulated in the dorsal subdomain of the Inner, Mid and Outer layers (Fig.23, indicated by *, Fig.24, Fig.25b), along with *FOXF1* in the outer domain. These TFs have been shown to be activated by sonic hedgehog (Shh) which may be found more highly in the dorsal domain due to its proximity to the notochord where *SHH* is expressed (Echelard *et al.*, 1993).

The area within the microdissected sub-domains from the CS17 analysis are roughly equivalent to the CS16 Inner + Mid layers combined. Spatially polarised genes were compared from the two analyses and there was little overlap which may represent differences between stages but as the library preparation method was different for each analysis this may also be a technical issue. When comparing the ventrally upregulated genes vs. the dorsal counterparts in both the CS17 and CS16 analyses, the CS17 ventral_all subdomain has 1 and 4 overlapping genes with the ventral Inner and Mid layers respectively (Fig.25a). *REN* was one of these genes and is one of the top ventrally enriched genes in both the CS17 dorsal aorta and CS16 Mid layer (log2FC=11.3, log2FC=5.8). *GATA3*, a factor known for promoting HSC emergence (Fitch *et al.*, 2012), was also enriched at both stages. The CS17 dorsal_all subdomain has only 3 and 4 overlapping genes with the dorsal Inner and Mid layers respectively despite the high number of upregulated genes (Fis.25b). This includes *FOXF2* further indicating that it may be an effector of Shh in the dorsal region of the dorsal aorta. As CS17 is at the end of the window of HSC and IAHC appearance in the AGM the different gene expression profiles between CS16 and CS17 may indicate a shift away from pro-haematopoietic signals.

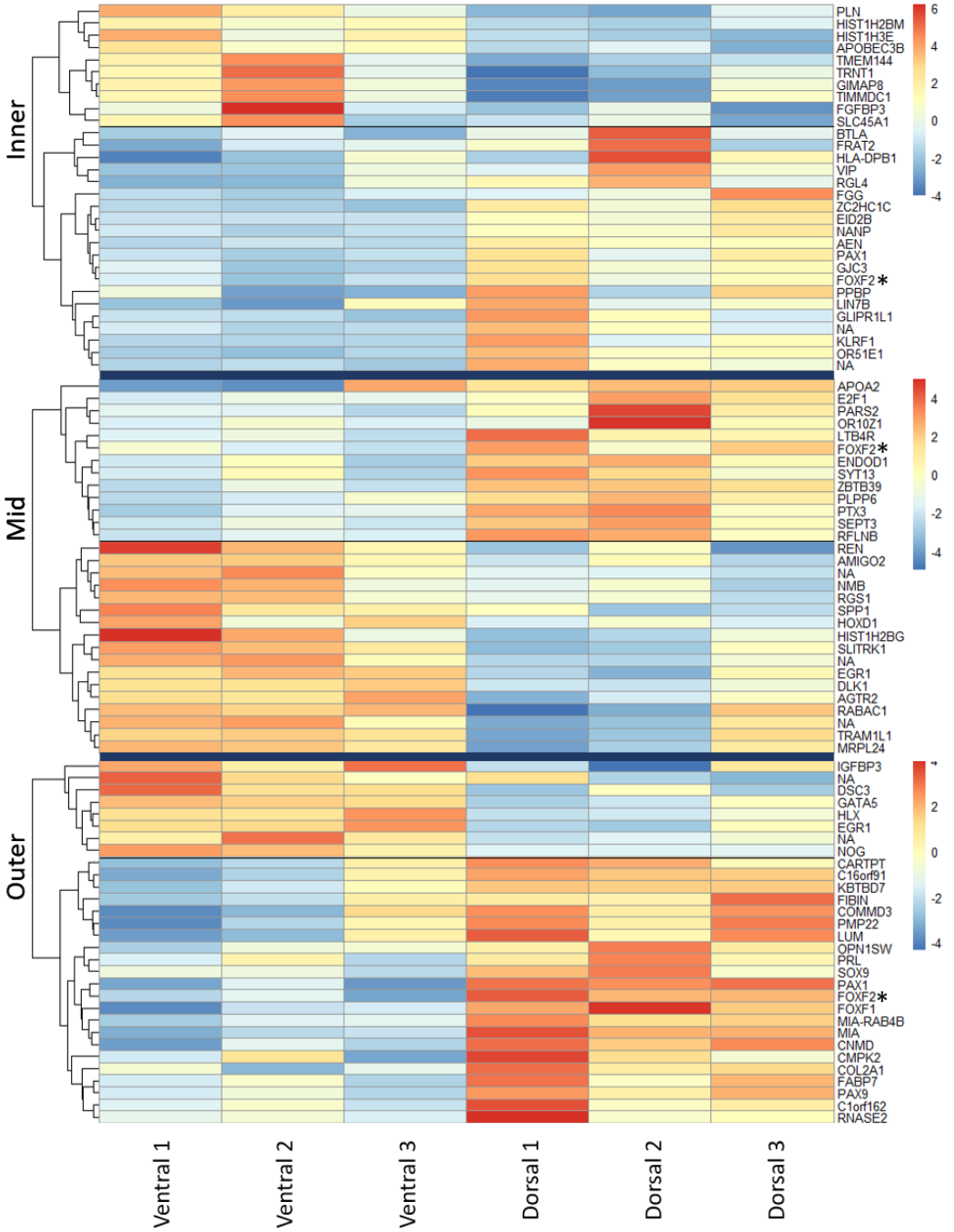


Figure 23. Polarised gene expression along the dorsal-ventral axis both close to the dorsal aorta and deeper into the underlying mesenchymal cells. Heatmap of top 30 differentially expressed genes between ventral and dorsal domains for each layer (Inner, Mid, Outer)(p.adj<0.05). NA = genes without an official gene name. * highlights FOXF2 which is dorsally enriched in each layer.

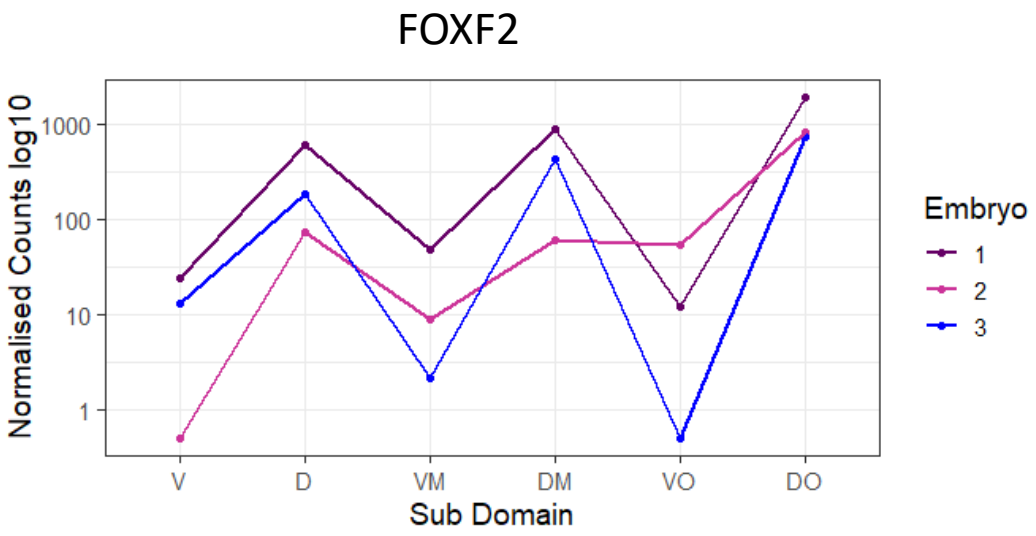
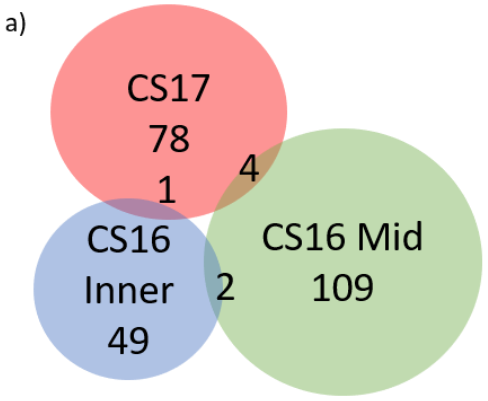
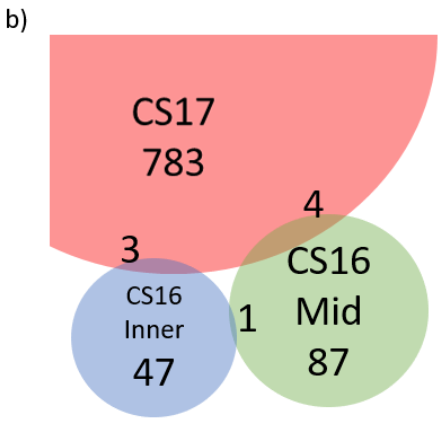


Figure 24. *FOXF2* is dorsally enriched in each concentric layer. Normalised log_{10} expression levels for FOXF2 across each sub-domain; V = Ventral Inner, D = Dorsal Inner, VM = Ventral Mid, DM = Dorsal Mid, VO = Ventral Outer, DO = Dorsal Outer





Ventral		
CS17 v CS16 Inner	**CS17 v CS16 Mid**	**CS16 Inner v Mid**
DIO3	REN	HIST1H3E
	EGR3	PLN
	BST2	
	GATA3	

Dorsal		
CS17 v CS16 Inner	**CS17 v CS16 Mid**	**CS16 Inner v Mid**
PAX1	PRND	FOXF2
FOXF2	FOXF2	
SOGA3	PTX3	
	ZBTB39	

Figure 25 : Polarised genes overlapping between CS17 and CS16 analyses. (a) shows the number of ventrally upregulated genes compared to the dorsal counterpart for CS17 ventral_all, CS16 Inner and CS16 Mid, and the number of genes overlapping between the analyses. (b) shows the number of dorsally upregulated genes compared to the ventral counterpart for CS17 dorsal_all, CS16 Inner and CS16 Mid, and the number of genes overlapping between the analyses. The tables below indicate the overlapping genes in both venn diagrams.

3.3.5 Gene Set Enrichment Analysis (GSEA) reveals both layer specific and dorsal-ventral polarised signalling pathways in the dorsal aorta

GSEA allows signalling pathways to be resolved to each concentric layer radiating away from the dorsal aorta as well as to determine D-V polarised signalling within each layer.

1) *Inner vs. Outer layers.* This analysis reveals signalling pathways enriched in the vessel wall which may be involved in such processes as angiogenesis, blood vessel wall maintenance or HSC generation (Fig. 26).
2) *Inner + Mid vs Outer layers.* Finding shared pathways enriched in the Inner and Mid layers compared with the Outer layer reveals cross talk between these layers (Fig.27).
3) *D-V Inner analysis.* This reveals polarised pathways across the D-V axis of the vessel wall which may influence polarised processes such as specification of the haematogenic endothelium or HSC/HSPC emergence (Fig.28).
4) *D-V Mid analysis.* This reveals polarised pathways across the D-V axis of the sub-aortic mesenchyme which may also influence polarised processes independently or through direct signalling interactions with the Inner layer (Fig.29).
5) *D-V Outer analysis.* This reveals polarised pathways across the D-V axis of the outer mesenchyme which, although at a greater distance from the site of HSC emergence, may influence this process and other polarised processes by means of secreted molecule diffusion, cell migration or signalling interactions with the Mid and Inner layers (Fig.30).

1) *Inner vs. Outer layers*

Pathways enriched for both the ventral and dorsal Inner layers (endothelial + perivascular cells) vs. the respective outer layers (Fig.26) reveals pathways enriched in the vessel wall that are not necessarily polarised. The V_Inner and D_Inner sub-domains had 270 and 227 pathways enriched when contrasted with the V_Outer and D_Outer respectively (Fig.26b). Of these pathways, 124 are the same and represent signalling pathways found along the length of the dorsal aorta D-V axis (Fig.26b, c). Top enriched pathways for this Inner layer include 'TNFA signalling via NFκB' (Hallmark), 'TGFβ signalling' (Hallmark) and 'Smooth Muscle Contraction' (Reactome) (Fig.26c).

2) *Inner + Mid vs. Outer layers*

The Inner and Mid layers can be considered to be the most important cell layers for supporting HSC generation and interactions between these two layers may be essential for this process. For this reason, it is useful to compare differentially enriched pathways for both the Inner and Mid layer with the Outer layer so that overlapping signalling pathways between the two can be established (Fig.27). In the ventral domain the V_Inner + V_Mid layers have 270 and 189 pathways enriched compared with

the V_Outer layer respectively (Fig.27a). Of these pathways, 105 were shared between V_Inner + V_Mid layers indicating a high degree of interactivity and possibly co-cooperativity of these layers with regard to processes occurring in the AGM region. In the dorsal domain compared to the Outer layer, the D_Inner layer has 227 pathways enriched whereas the D_Mid layer has only 33 (Fig.27b). Of these 33, 27 are shared with the D_Inner layer indicating less signalling interactions between D_Inner + D_Mid layers than in the ventral domain and greater Inner layer specific signalling in the dorsal domain. Indeed, pathways from DE genes between D_Inner + D_Mid layers show 68 Inner pathways and 14 Mid pathways. The same analysis in the ventral domain show much fewer differentially enriched pathways with 12 and 24 pathways enriched for V_Inner + V_Mid respectively.

Top pathways for the overlap of Inner + Mid vs Outer in both dorsal and ventral domains included 'TGFβ signalling' (Hallmark), 'Epithelial-Mesenchymal Transition' (Hallmark), 'TNFA signalling via NFκB' (Hallmark), and 'Smooth Muscle Contraction' (Reactome) indicating that these signalling pathways are required across the full D-V axis of the aorta (Fig.27a,b). There are however, D-V differences, and whether these are significant or not was determined through D-V differential gene expression analysis of each layer and subsequent GSEA.

3) D-V Inner analysis

GSEA of V_Inner vs D_Inner differentially expressed genes revealed polarised signalling pathways with 10 pathways enriched in the ventral domain and 22 enriched in the dorsal (Fig.28). Interestingly, 'TGFβ pathway' (Hallmark, Biocarta) and 'TNFA signalling via NFκB' (Hallmark), are ventrally enriched despite being enriched in both dorsal and ventral domains when contrasted with the outer layer. This demonstrates a gradient of these pathways towards the ventral domain suggesting that they are more active there. Conversely, 'Angiogenesis' (Hallmark), 'IL6 JAK STAT3 Signalling' (Hallmark), 'Formation of Fibrin Clot Clotting Cascade' (Reactome), 'AMI Pathway' (Acute Myocardial Infarction pathway, Biocarta), 'Inflammatory Response' (Hallmark) and 'Coagulation' (Hallmark) are found in both domains of the Inner layer vs . Outer layer but are dorsally enriched revealing a gradient of signalling towards the dorsal domain. Gene sets to do with transcription are ventrally enriched including 'RNA Pol I Transcription' (Reactome), and 'RNA Pol I Promoter Opening' (Reactome). Strangely, 'Meiosis' (Reactome) is also enriched in the ventral domain but a closer look at the genes contributing to this pathway enrichment reveals they are mainly encoding histone subunits and not meiosis specific proteins. This gives further evidence for higher transcriptional activity in the V_Inner domain. In the dorsal domain such inflammation pathways as 'Inflammatory Response' (Hallmark), 'Cytokine-Cytokine Receptor Interaction' (KEGG), 'IL6 JAK STAT3 Signalling' (Hallmark) and 'NKT Pathway' (Selective expression of chemokine receptors during T-cell polarization, Biocarta) are enriched.

4) *D-V Mid Analysis*

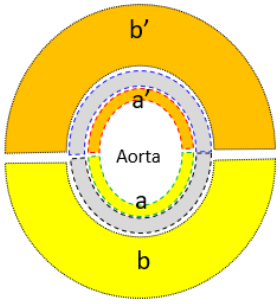
In order to investigate whether signalling is polarised in the sub-aortic mesenchyme GSEA of the DE genes between V_Mid and D_Mid layers was carried out and revealed many spatially polarised signalling pathways (Fig.29). In the ventral domain, 'TGFβ pathway' (Hallmark, Biocarta) and 'TNFA signalling via NFκB' (Hallmark) are again enriched. The top (lowest FDR) ventral pathway is 'Oxidative Phosphorylation' suggesting cells are more metabolically active in this domain. 'Protein Secretion' (Hallmark) is also ventrally enriched which may indicate there is more inter-cellular communication in this domain. Other notable ventral pathways include pathways regulated by the oncogene MYC; 'MYC Targets V1' (Hallmark) and tumor suppressor P53; 'P53 Pathway' (Hallmark), indicative of controlled cell proliferation in this domain. In the dorsal domain the 'Olfactory Signalling Pathway' (Reactome) is the top enriched pathway which was also enriched in the dorsal Inner domain (Fig.28). 'KRAS Signalling DN' (Genes downregulated by KRAS activation, Hallmark) was also dorsally enriched in both Inner and Mid domains suggesting k-RAS could be polarising gene expression in the dorsal aorta.

5) *D-V Outer Analysis*

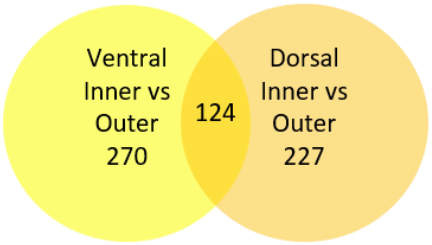
GSEA for DE genes between V_Outer and D_Outer subdomains revealed ventral enrichment of several pathways associated with haematopoietic factors such as TPO ('TPO Pathway', Biocarta), EPO ('EPO Pathway', PID and Biocarta), PDGF ('PDGF Pathway', Biocarta and 'PDGFRA Pathway', PID) and IL3 ('IL3 Pathways', Biocarta) (Fig.30). It is conceivable that ventral layers distal to the dorsal aorta can contribute to the haematopoietic niche by emanating long-range signals. In the dorsal domain, 26 out of 103 enriched pathways were directly related to regulation of cell cycle and mitosis indicating high levels of proliferation in this domain. There are several enriched pathways involved in cell metabolism and transcription but these are cell processes rather than classical signalling pathways suggesting that cell-cell communication is reduced in the dorsal domain compared to the ventral.

This analysis demonstrates that there are both non-polarised and D-V polarised signalling pathways enriched in the dorsal aorta vessel wall. The high number of overlapping enriched pathways between V_Inner and V_Mid, indicates high levels of crosstalk between these two layers. The signalling in these layers close to the dorsal aorta are quite distinct from V/D_Outer pathways. Nonetheless, the Outer layer demonstrated D-V polarised pathways including ventrally enriched pathways involving haematopoietic factors. Taken together, this suggests that cells both proximal and distal to the dorsal aorta may generate niche signals influencing HSC emergence.

a)



b)

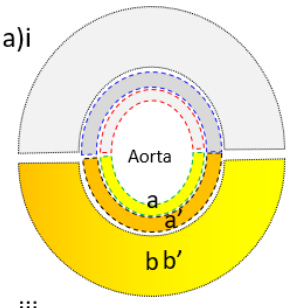


c)

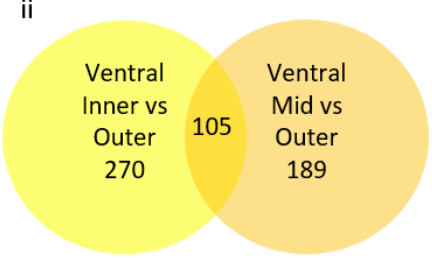
Inner v Outer GSEA - Ventral/Dorsal Overlap		
	NES	FDR
HALLMARK_TNFA_SIGNALING_VIA_NFKB	2.180133	0
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	2.140082	0
HALLMARK_TGF_BETA_SIGNALING	2.108407	0
REACTOME_SMOOTH_MUSCLE_CONTRACTION	2.37653	0
PID_S1P_S1P1_PATHWAY	2.225124	0
HALLMARK_HYPOXIA	1.930139	5.59E-04
HALLMARK_IL2_STAT5_SIGNALING	1.792559	0.004829
HALLMARK_INTERFERON_GAMMA_RESPONSE	1.765629	0.005397
HALLMARK_MTORC1_SIGNALING	1.642325	0.011949
HALLMARK_WNT_BETA_CATENIN_SIGNALING	1.579654	0.01567
HALLMARK_UV_RESPONSE_DN	1.534534	0.015755
HALLMARK_APOPTOSIS	1.599266	0.016309
HALLMARK_COAGULATION	1.539908	0.016673
HALLMARK_ANDROGEN_RESPONSE	1.552653	0.016688
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	1.562074	0.016777
HALLMARK_MYOGENESIS	1.579974	0.017412
PID_PDGFRB_PATHWAY	2.04265	0.020461
HALLMARK_PI3K_AKT_MTOR_SIGNALING	1.498512	0.020902
HALLMARK_APICAL_JUNCTION	1.459742	0.023995
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	2.044505	0.024643
HALLMARK_REACTIVE_OXIGEN_SPECIES_PATHWAY	1.447787	0.024804
PID_NECTIN_PATHWAY	1.944407	0.024852
REACTOME_MUSCLE_CONTRACTION	1.992123	0.02492
BIOCARTA_RHO_PATHWAY	1.938539	0.025042
HALLMARK_NOTCH_SIGNALING	1.440861	0.025227
REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS	1.980205	0.026356

Figure 26 Top pathways enriched in the Inner layer compared to the Outer layer reveals signalling of the dorsal aorta vessel wall. a) Schematic of contrast for GSEA a vs. b and a' vs. b'. b) Number of pathways enriched in the Ventral Inner vs Ventral Outer and the Dorsal Inner vs Dorsal Outer and the number of shared pathways. c) Overlap between enriched pathways in the ventral and dorsal Inner domains compared to the respective Outer domain. Top 25 are shown. FDR < 0.25, NES = Normalised Enrichment Score.

a)i



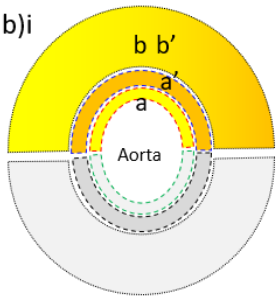
ii



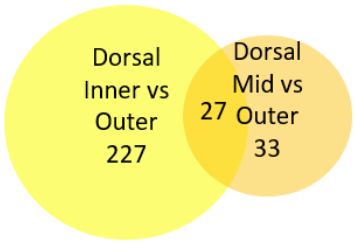
iii

Ventral Inner and Mid vs Outer GSEA - Inner/Mid Overlap	Inner		Mid	
	NES	FDR	NES	FDR
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	2.140082	0	2.299595	0
HALLMARK_TGF_BETA_SIGNALING	2.108407	0	1.671448	0.006854
HALLMARK_TNFA_SIGNALING_VIA_NFKB	2.180133	0	2.072549	0
REACTOME_SMOOTH_MUSCLE_CONTRACTION	2.37653	0	2.624209	0
HALLMARK_HYPOXIA	1.930139	5.59E-04	1.685197	0.007347
HALLMARK_IL2_STAT5_SIGNALING	1.792559	0.004829	1.152416	0.202269
HALLMARK_INTERFERON_GAMMA_RESPONSE	1.765629	0.005397	1.244772	0.104107
HALLMARK_MTORC1_SIGNALING	1.642325	0.011949	1.622641	0.013902
HALLMARK_WNT_BETA_CATENIN_SIGNALING	1.579654	0.01567	1.415736	0.035621
HALLMARK_UV_RESPONSE_DN	1.534534	0.015755	1.34855	0.053103
HALLMARK_APOPTOSIS	1.599266	0.016309	1.625866	0.015446
HALLMARK_COAGULATION	1.539908	0.016673	1.359699	0.049388
HALLMARK_ANDROGEN_RESPONSE	1.552653	0.016688	1.363535	0.051133
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	1.562074	0.016777	1.859873	0.001284
HALLMARK_ADIPOGENESIS	1.522453	0.017161	1.519826	0.020656
HALLMARK_MYOGENESIS	1.579974	0.017412	1.579325	0.014569
HALLMARK_OXIDATIVE_PHOSPHORYLATION	1.513345	0.01772	2.510108	0
PID_PDGFRB_PATHWAY	2.04265	0.020461	1.626511	0.090713
HALLMARK_MYC_TARGETS_V1	1.479333	0.022685	2.418597	0
HALLMARK_APICAL_JUNCTION	1.459742	0.023995	1.194562	0.143529

b)i



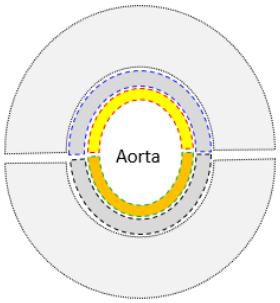
ii



iii

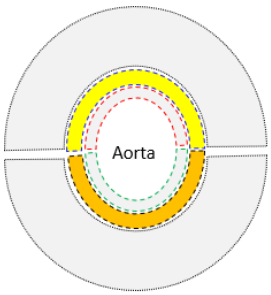
Dorsal Inner and Mid vs Outer GSEA - Inner/Mid Overlap	Inner		Mid	
	NES	FDR	NES	FDR
HALLMARK_TGF_BETA_SIGNALING	2.02281	0	1.636286	0.021287
HALLMARK_TNFA_SIGNALING_VIA_NFKB	2.148233	0	1.489274	0.043148
REACTOME_MUSCLE_CONTRACTION	2.496085	0	2.177408	0.003505
REACTOME_SMOOTH_MUSCLE_CONTRACTION	2.568864	0	2.327986	0
HALLMARK_ANGIOGENESIS	1.910155	0.000707	1.48298	0.037845
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	1.851159	9.79E-04	1.818135	0.008804
HALLMARK_IL2_STAT5_SIGNALING	1.788941	0.002727	1.492524	0.052952
HALLMARK_APICAL_JUNCTION	1.672924	0.006525	1.334673	0.094293
HALLMARK_P53_PATHWAY	1.566487	0.012924	1.323035	0.095403
HALLMARK_MYOGENESIS	1.579421	0.013221	1.40695	0.059477
HALLMARK_APOPTOSIS	1.586134	0.013481	1.42675	0.056526
PID_EPO_PATHWAY	1.949149	0.01787	1.696385	0.245902
HALLMARK_WNT_BETA_CATENIN_SIGNALING	1.461683	0.028977	1.538292	0.043701
PID_S1P_S1P1_PATHWAY	1.860937	0.034114	1.684596	0.231082
PID_EPHA_FWDPATHWAY	1.814386	0.035554	1.727471	0.192523
PID_AMB2_NEUTROPHILS_PATHWAY	1.80594	0.038083	1.847181	0.101423
REACTOME_PEPTIDE_CHAIN_ELONGATION	1.780073	0.040947	2.156592	0.003253
PID_LYMPH_ANGIOGENESIS_PATHWAY	1.760965	0.046446	1.78126	0.141163
KEGG_VASCULAR_SMOOTH_MUSCLE_CONTRACTION	1.702746	0.065783	2.070646	0.006925
REACTOME_INFLUENZA_VIRAL_RNA_TRANSCRIPTION_AND_REPLICATION	1.703492	0.066457	1.928617	0.041928

Figure 27 (Previous page) Overlapping enriched pathways in the Inner and Mid domains vs. the outer domain highlight shared signalling pathways. i. Schematic of contrast for GSEA a vs. b and a' vs. b' for ventral a) and dorsal b) ii. Number of pathways enriched in Inner vs. Outer and Mid vs. Outer. iii. Top 20 (by Inner FDR) overlapping Inner + Mid vs. Outer pathways. FDR < 0.25, NES = Normalised Enrichment Score. Overlapping pathways between ventral (a) and dorsal (b) are colour coded in blue.



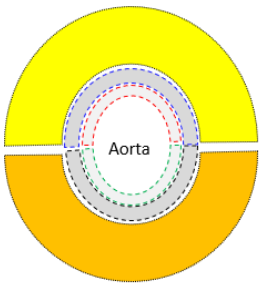
Inner Ventral vs Dorsal GSEA					
Ventral	NES	FDR	**Dorsal**	NES	FDR
REACTOME_MEIOSIS	1.92	0.08	HALLMARK_KRAS_SIGNALING_DN	-1.91	0.00
BIOCARTA_TGFB_PATHWAY	1.96	0.08	REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES	-2.17	0.00
REACTOME_MEIOTIC_RECOMBINATION	1.98	0.09	REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS	-2.10	0.00
KEGG_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	1.93	0.10	KEGG_OLFACTORY_TRANSDUCTION	-1.97	0.01
REACTOME_RNA_POL_I_RNA_POL_III_AND_MITOCHONDRIAL_TRANSCRIPTION	1.90	0.10	REACTOME_OLFACTORY_SIGNALING_PATHWAY	-1.94	0.01
REACTOME_RNA_POL_I_TRANSCRIPTION	2.00	0.10	HALLMARK_ANGIOGENESIS	-1.66	0.02
HALLMARK_TGF_BETA_SIGNALING	1.53	0.13	HALLMARK_IL6_JAK_STAT3_SIGNALING	-1.47	0.09
REACTOME_SYNTHESIS_OF_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI	1.83	0.17	HALLMARK_ALLOGRAFT_REJECTION	-1.48	0.10
REACTOME_RNA_POL_I_PROMOTER_OPENING	2.02	0.17	REACTOME_G_ALPHA_I_SIGNALLING_EVENTS	-1.78	0.10
HALLMARK_TNFA_SIGNALING_VIA_NFKB	1.40	0.23	REACTOME_GPCR_LIGAND_BINDING	-1.78	0.10
			REACTOME_FORMATION_OF_FIBRIN_CLOT_CLOTTING_CASCADE	-1.79	0.10
			REACTOME_CLASS_A1_RHODOPSIN_LIKE_RECEPTORS	-1.81	0.11
			BIOCARTA_AMI_PATHWAY	-1.76	0.11
			KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	-1.74	0.12
			HALLMARK_INFLAMMATORY_RESPONSE	-1.36	0.16
			HALLMARK_COAGULATION	-1.38	0.17
			KEGG_DRUG_METABOLISM_CYTOCHROME_P450	-1.70	0.18
			NABA_ECM_REGULATORS	-1.68	0.20
			KEGG_AUTOIMMUNE_THYROID_DISEASE	-1.65	0.22
			BIOCARTA_NKT_PATHWAY	-1.65	0.22
			REACTOME_PHASE_II_CONJUGATION	-1.66	0.22
			KEGG_ALLOGRAFT_REJECTION	-1.66	0.24

Figure 28 GSEA of differentially expressed genes between the ventral and dorsal sub-domains of the Inner layer. FDR < 0.25, NES = Normalised Enrichment Score.



Mid Ventral vs Dorsal GSEA					
Ventral	**NES**	**FDR**	**Dorsal**	**NES**	**FDR**
HALLMARK_OXIDATIVE_PHOSPHORYLATION	2.2401	0	REACTOME_OLFACTORY_SIGNALING_PATHWAY	-2.49	0
HALLMARK_TNFA_SIGNALING_VIA_NFKB	1.8992	0.002	KEGG_OLFACTORY_TRANSDUCTION	-2.442	0
HALLMARK_MYC_TARGETS_V1	1.9127	0.003	HALLMARK_KRAS_SIGNALING_DN	-1.813	0.001
HALLMARK_PROTEIN_SECRETION	1.8246	0.004	KEGG_AUTOIMMUNE_THYROID_DISEASE	-2.019	0.003
REACTOME_ENDOSOMAL_SORTING_COMPLEX_REQUIRED_FOR_TRANSPORT_ESCRT	2.2568	0.006	KEGG_ASTHMA	-1.952	0.01
HALLMARK_DNA_REPAIR	1.6671	0.011	REACTOME_BETA_DEFENSINS	-1.908	0.017
HALLMARK_TGF_BETA_SIGNALING	1.578	0.027	KEGG_LINOLEIC_ACID_METABOLISM	-1.894	0.018
HALLMARK_ANDROGEN_RESPONSE	1.4489	0.037	REACTOME_DEFENSINS	-1.857	0.025
HALLMARK_MTORC1_SIGNALING	1.5107	0.038	KEGG_STEROID_HORMONE_BIOSYNTHESIS	-1.836	0.027
HALLMARK_HYPOXIA	1.5328	0.038	BIOCARTA_CYTOKINE_PATHWAY	-1.858	0.028
HALLMARK_UV_RESPONSE_UP	1.4498	0.04	KEGG_TASTE_TRANSDUCTION	-1.837	0.029
HALLMARK_UV_RESPONSE_DN	1.4565	0.041	REACTOME_PHASE1_FUNCTIONALIZATION_OF_COMPOUNDS	-1.748	0.082
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	1.4622	0.042	REACTOME_CYTOCHROME_P450_ARRANGED_BY_SUBSTRATE_TYPE	-1.723	0.095
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	1.4753	0.042	KEGG_ARACHIDONIC_ACID_METABOLISM	-1.724	0.103
HALLMARK_P53_PATHWAY	1.4243	0.043	NABA_SECRETED_FACTORS	-1.704	0.114
HALLMARK_CHOLESTEROL_HOMEOSTASIS	1.4137	0.045	REACTOME_ACYL_CHAIN_REMODELLING_OF_PS	-1.679	0.128
REACTOME_AMYLOIDS	2.0586	0.056	KEGG_PHENYLALANINE_METABOLISM	-1.666	0.133
REACTOME_RNA_POL_I_PROMOTER_OPENING	2.005	0.061	PID_CONE_PATHWAY	-1.686	0.133
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS_	2.0935	0.062	KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	-1.67	0.134
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	2.0215	0.063	KEGG_ALLOGRAFT_REJECTION	-1.679	0.135
HALLMARK_COAGULATION	1.3428	0.072	PID_IL23_PATHWAY	-1.66	0.136
ST_GRANULE_CELL_SURVIVAL_PATHWAY	1.8586	0.108	KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	-1.636	0.169
REACTOME_LYSOSOME_VESICLE_BIOGENESIS	1.8989	0.109	REACTOME_AMINE_DERIVED_HORMONES	-1.623	0.177
REACTOME_CHOLESTEROL_BIOSYNTHESIS	1.8612	0.115	REACTOME_COSTIMULATION_BY_THE_CD28_FAMILY	-1.571	0.182
REACTOME_MEIOTIC_RECOMBINATION	1.8708	0.115	REACTOME_BIOLOGICAL_OXIDATIONS	-1.576	0.183
BIOCARTA_EIF_PATHWAY	1.8067	0.122	KEGG_GRAFT_VERSUS_HOST_DISEASE	-1.624	0.183
REACTOME_MEMBRANE_TRAFFICKING	1.9005	0.123	PID_RHODOPSIN_PATHWAY	-1.616	0.183
REACTOME_MITOCHONDRIAL_PROTEIN_IMPORT	1.8737	0.124	PID_CD40_PATHWAY	-1.572	0.184
PID_SMAD2_3PATHWAY	1.811	0.124	BIOCARTA_DC_PATHWAY	-1.588	0.187
REACTOME_PACKAGING_OF_TELOMERE_ENDS	1.7922	0.128	KEGG_MATURITY_ONSET_DIABETES_OF_THE_YOUNG	-1.576	0.187
REACTOME_BIOSYNTHESIS_OF_THE_N_GLYCAN_PRECURSOR_DOLICHOL_LIPID_LINKED_OLIGOSACCHARIDE_LLO_AND_TRANSFER_TO_A_NASCENT_PROTEIN	1.9091	0.131	REACTOME_GLUCURONIDATION	-1.593	0.191
PID_FRA_PATHWAY	1.8112	0.133	REACTOME_AMINE_LIGAND_BINDING_RECEPTORS	-1.608	0.191
BIOCARTA_CCR5_PATHWAY	1.8201	0.133	REACTOME_TRAF6_MEDIATED_IRF7_ACTIVATION	-1.577	0.192
HALLMARK_MITOTIC_SPINDLE	1.2082	0.175	KEGG_TYPE_I_DIABETES_MELLITUS	-1.56	0.192
HALLMARK_GLYCOLYSIS	1.216	0.176	REACTOME_ACYL_CHAIN_REMODELLING_OF_PG	-1.589	0.193
HALLMARK_APOPTOSIS	1.1948	0.185	NABA_PROTEOGLYCANS	-1.556	0.194
PID_PRL_SIGNALING_EVENTS_PATHWAY	1.7285	0.195	KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	-1.595	0.194
PID_INTEGRIN_A4B1_PATHWAY	1.7349	0.195	REACTOME_ACYL_CHAIN_REMODELLING_OF_PE	-1.561	0.195
REACTOME_CITRIC_ACID_CYCLE_TCA_CYCLE	1.7072	0.219	KEGG_DRUG_METABOLISM_CYTOCHROME_P450	-1.578	0.195
BIOCARTA_CARM_ER_PATHWAY	1.6836	0.249	REACTOME_AMINO_ACID_TRANSPORT_ACROSS_THE_PLASMA_MEMBRANE	-1.598	0.196
			REACTOME_XENOBIOTICS	-1.58	0.197
			REACTOME_G_ALPHA_S_SIGNALLING_EVENTS	-1.549	0.202
			BIOCARTA_SHH_PATHWAY	-1.599	0.202
			REACTOME_PD1_SIGNALING	-1.542	0.204
			KEGG_ALPHA_LINOLENIC_ACID_METABOLISM	-1.542	0.207
			KEGG_PRIMARY_BILE_ACID_BIOSYNTHESIS	-1.543	0.21
			REACTOME_INTERACTION_BETWEEN_L1_AND_ANKYRINS	-1.526	0.221
			KEGG_RETINOL_METABOLISM	-1.528	0.221
			REACTOME_AMINO_ACID_AND_OLIGOPEPTIDE_SLC_TRANSPORTERS	-1.523	0.221
			REACTOME_POTASSIUM_CHANNELS	-1.53	0.221

Figure 29. GSEA of differentially expressed genes between the ventral and dorsal sub-domains of the Mid layer. FDR < 0.25, NES = Normalised Enrichment Score.



Outer Ventral vs Dorsal GSEA					
Ventral	NES	FDR	Dorsal	NES	FDR
PID_NFAT_TFPATHWAY	2.13218	0.019698	HALLMARK_E2F_TARGETS	-2.25453	0
ST_DIFFERENTIATION_PATHWAY_IN_PC12_CELLS	2.011431	0.03523	HALLMARK_G2M_CHECKPOINT	-2.06058	2.46E-04
REACTOME_NUCLEOTIDE_LIKE_PURINERGIC_RECEPTORS	2.024404	0.039453	HALLMARK_MYC_TARGETS_V1	-2.15938	3.68E-04
PID_PDGFRA_PATHWAY	2.046053	0.045223	HALLMARK_DNA_REPAIR	-1.87903	8.13E-04
PID_CD8_TCR_DOWNSTREAM_PATHWAY	1.935155	0.075681	REACTOME_DNA_REPLICATION	-2.18712	0.001421
BIOCARTA_IGF1_PATHWAY	1.775494	0.179924	HALLMARK_OXIDATIVE_PHOSPHORYLATION	-1.70257	0.003624
BIOCARTA_IL3_PATHWAY	1.782717	0.181579	HALLMARK_FATTY_ACID_METABOLISM	-1.63175	0.009311
REACTOME_REGULATION_OF_INSULIN_LIKE_GROWTH_FACTOR_IGF_ACTIVITY_BY_INSULIN_LIKE_GROWTH_FACTOR_BINDING_PROTEINS_IGFBPS	1.795218	0.182392	KEGG_DNA_REPLICATION	-2.00376	0.010732
PID_EPO_PATHWAY	1.831537	0.189228	REACTOME_EXTENSION_OF_TELOMERES	-1.99605	0.011103
PID_TCR_CALCIUM_PATHWAY	1.757543	0.198762	REACTOME_SYNTHESIS_OF_DNA	-1.98006	0.011168
REACTOME_IL_3_5_AND_GM_CSF_SIGNALING	1.810131	0.201931	REACTOME_CELL_CYCLE_MITOTIC	-2.03112	0.012828
BIOCARTA_TPO_PATHWAY	1.797193	0.202754	REACTOME_MITOTIC_M_M_G1_PHASES	-2.0042	0.013234
REACTOME_REGULATION_OF_SIGNALING_BY_CBL	1.736414	0.220179	REACTOME_GLUCONEOGENESIS	-2.05768	0.014442
BIOCARTA_PDGF_PATHWAY	1.692645	0.223948	REACTOME_S_PHASE	-1.92441	0.023644
PID_NFKAPPAB_ATYPICAL_PATHWAY	1.684813	0.228716	REACTOME_G1_S_TRANSITION	-1.9112	0.024405
PID_P38_ALPHA_BETA_PATHWAY	1.695056	0.231223	HALLMARK_MYC_TARGETS_V2	-1.52022	0.028409
BIOCARTA_IL2_PATHWAY	1.705581	0.236407	REACTOME_NONSENSE_MEDIATED_DECAY_ENHANCED_BY_THE_EXON_JUNCTION_COMPLEX	-1.8656	0.030611
BIOCARTA_EPO_PATHWAY	1.698572	0.237679	KEGG_RIBOSOME	-1.84947	0.030677
BIOCARTA_EGF_PATHWAY	1.709196	0.244179	KEGG_PYRIMIDINE_METABOLISM	-1.87013	0.031235
BIOCARTA_TCR_PATHWAY	1.709945	0.260416	REACTOME_APC_C_CDC20_MEDIATED_DEGRADATION_OF_MITOTIC_PROTEINS	-1.85182	0.031558
			REACTOME_CELL_CYCLE_CHECKPOINTS	-1.88301	0.031677
			PID_AURORA_B_PATHWAY	-1.85401	0.032729
			REACTOME_REGULATION_OF_MITOTIC_CELL_CYCLE	-1.87015	0.034074
			REACTOME_LAGGING_STRAND_SYNTHESIS	-1.8322	0.034082
			REACTOME_SCFSKP2_MEDIATED_DEGRADATION_OF_P27_P21	-1.82489	0.035402
			REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	-1.83267	0.035957
			REACTOME_G1_S_SPECIFIC_TRANSCRIPTION	-1.81106	0.036248
			REACTOME_MITOTIC_G1_G1_S_PHASES	-1.81797	0.036603
			REACTOME_ORC1_REMOVAL_FROM_CHROMATIN	-1.81424	0.036631
			KEGG_CITRATE_CYCLE_TCA_CYCLE	-1.79006	0.040976
			REACTOME_MITOTIC_PROMETAPHASE	-1.79069	0.042376
			REACTOME_CELL_CYCLE	-1.79377	0.042467
			KEGG_PYRUVATE_METABOLISM	-1.78309	0.042603
			REACTOME_APC_C_CDH1_MEDIATED_DEGRADATION_OF_CDC20_AND_OTHER_APC_C_CDH1_TARGETED_PROTEINS_IN_LATE_MITOSIS_EARLY_G1	-1.76635	0.049032
			BIOCARTA_MCM_PATHWAY	-1.75436	0.051995
			REACTOME_M_G1_TRANSITION	-1.75494	0.053466
			REACTOME_METABOLISM_OF_RNA	-1.747	0.053699
			BIOCARTA_PROTEASOME_PATHWAY	-1.73994	0.056079

Figure 30 GSEA of differentially expressed genes between the ventral and dorsal sub-domains of the Outer layer. FDR < 0.25, NES = Normalised Enrichment Score.

3.3.6 Gradients of gene expression towards and away from the dorsal aorta

To understand how signals in the AGM niche may be contributing to HSC generation, it is useful to know which expression of signalling molecules grow in intensity towards the dorsal aorta, the site of HSC emergence. For this, DESeq2 was used with a likelihood ratio test (LRT) as before.

From the top 100 most significant genes, secreted molecules were identified by mapping Gene IDs with the Uniprot keyword Secreted [KW-0964]). This gave 23 secreted proteins, 10 of which were extracellular matrix (ECM) proteins (Fig.31a, bii) demonstrating that the composition of the ECM changes towards the dorsal aorta. The most differentially expressed secreted protein was *EDN1*, a vasoconstrictor protein (Fig.30a, bi) which is also significantly ventrally enriched in the Mid layer (p.adj<0.05).

a

Gene Name	log2FC	p.adj
EDN1	6.61	2E-07
ANXA1	6.52	1E-07
HAPLN3	5.95	2E-10
LEFTY2	5.59	4E-04
CTGF	5.06	1E-04
IL17B	4.93	9E-05
TFPI	4.76	4E-15
BGN	4.64	2E-08
THBS1	4.57	1E-06
MGP	4.57	3E-09
VEGFC	4.39	2E-07
FLT1	4.28	9E-11
COL4A1	4.26	3E-10
ELN	4.24	2E-05
TEK	4.21	9E-12
INHBA	4.06	3E-09
CYR61	3.69	8E-04
HSPG2	3.67	1E-11
ACAN	3.48	7E-04
LAMA4	3.07	1E-04
CRIM1	2.93	2E-03
HTRA1	2.91	1E-03
FN1	2.38	1E-03

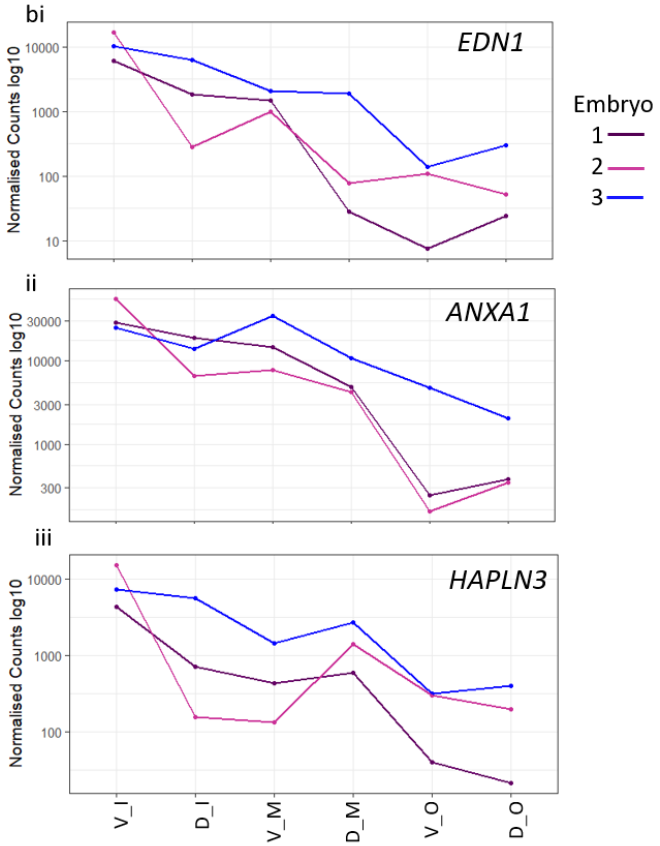
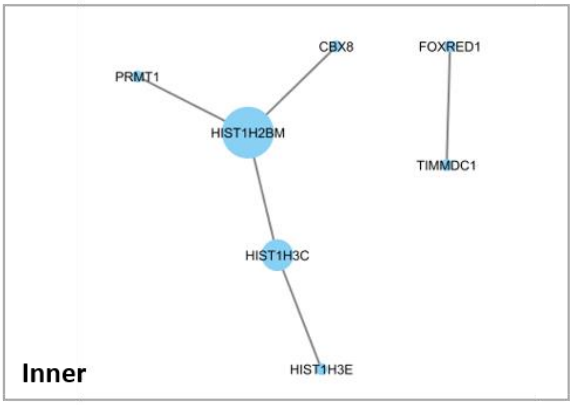
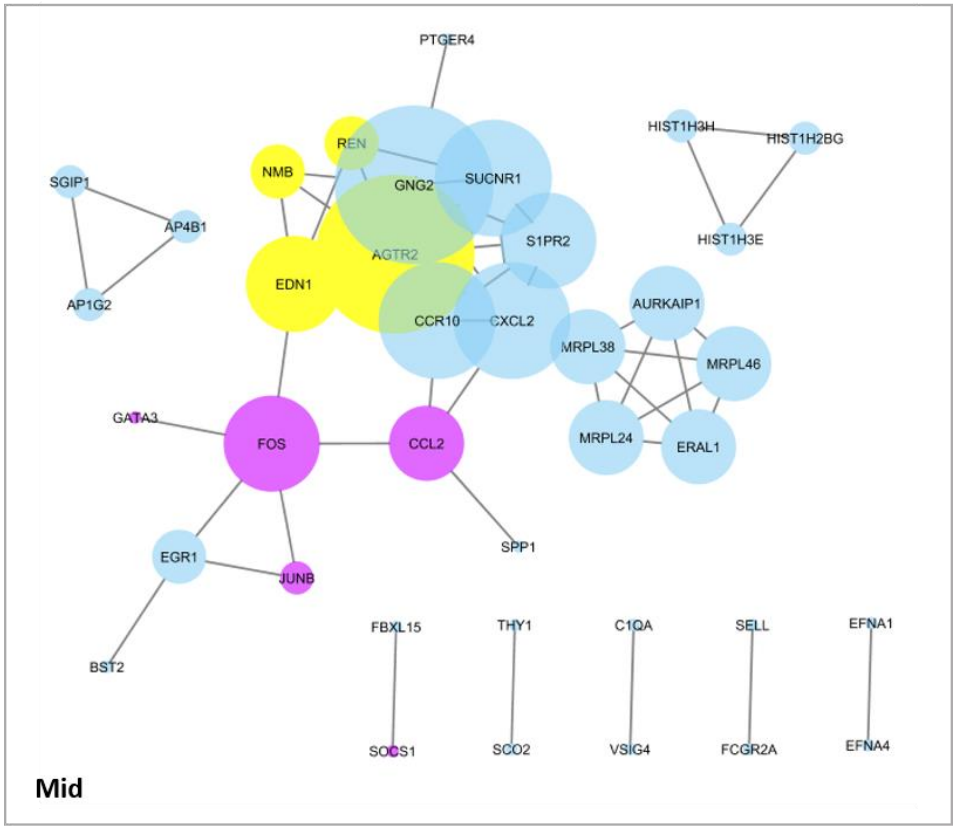


Figure 31 : Gradients of gene expression towards the dorsal aorta. (a) Secreted factors with gene expression increasing towards the dorsal aorta (Outer→Mid→ Inner) (Uniprot Keyword Secreted [KW-0964]). b) Normalised $\log_{10}$ counts of top 3 secreted proteins by $\log_2$ Fold Change.

3.3.7 Blood pressure regulatory genes are ventrally enriched

The STRING database (Szklarczyk *et al.*, 2019) was used to investigate interactions between the proteins encoded by significantly upregulated genes in the ventral domains of each layer with the aim of elucidating central players in the signalling networks (Fig.32).


CBX8
FOXRED1
PRMT1
HIST1H2BM
TIMMDC1
HIST1H3C
HIST1H3E
Inner


PTGER4
REN
HIST1H3H
HIST1H2BG
NMB
SUCNR1
GNG2
SGIP1
AP4B1
S1PR2
HIST1H3E
AGTR2
EDN1
AP1G2
AURKAIP1
CCR10
CXCL2
MRPL38
MRPL46
GATA3
FOS
CCL2
MRPL24
ERAL1
SPP1
EGR1
JUNB
BST2
FBXL15
THY1
C1QA
SELL
EFNA1
SOCS1
SCO2
VSIG4
FCGR2A
EFNA4
Mid

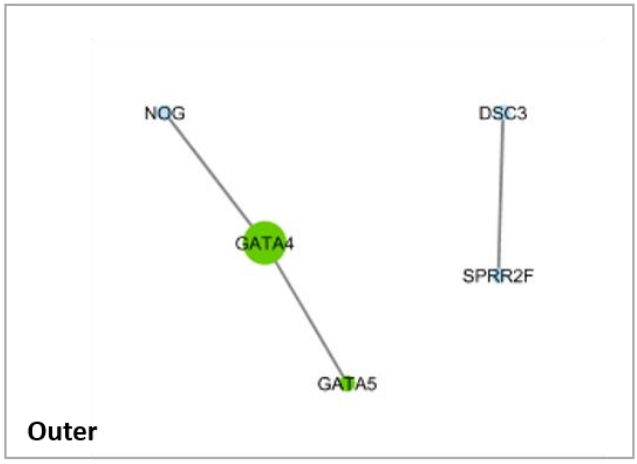

NOG
DSC3
GATA4
SPRR2F
GATA5
Outer

Figure 32 (previous page). A protein interactome model of the CS16 LCM-Seq ventral Inner, Mid and Outer layers. StringDB high confidence (0.7) interactions between significant (p.adj<0.05) ventrally enriched genes for each layer. Only nodes with connections to at least one other node are shown. Size of node increases with number of connecting edges to other nodes. Mid grid; Purple indicates "Interleukin-4 and Interleukin-13 signalling pathway", (Reactome) components. Yellow nodes indicate proteins involved in regulation of blood pressure regulation. Outer grid; green nodes show GATA zinc finger as annotated by PFAM Protein Domains database (FDR 0.00069)

Inner layer. In the Inner layer, out of 52 enriched genes, 7 showed high confidence interactions (Interaction score ≥ 0.7) (Fig.32a). 3 of these were encoding histone proteins (*HIST1H2BM, HIST1H3E, HIST1H3C).* The presence of histone encoding genes suggests that new histones are required for correct packaging of newly replicated DNA therefore suggesting there is increased cell division in the V_Inner compared to D_Inner. There are no upregulated histone genes in D_Inner (Fig.32) *HIST1H2BM* had 2 further connections to the epigenetic regulators *PRMT1* and *CBX8*.

Mid layer. In the Mid layer, due to the high number of significant genes only the top 100 were analysed. This analysis yielded many more interacting genes than the Inner layer (Fig.32). Notably, genes involved in blood pressure regulation and/or smooth muscle contraction were present and highly connected to each other including components of the Renin-Angiotensin system *REN* and *AGTR2* as well as *NMB* and *EDN1.* Components of the Reactome pathway 'Interleukin-4 and Interleukin-13 signalling pathway' were enriched within this string network including activator protein (AP-1) TFs *FOS* and *JUNB* as well as *CCL2,* and *GATA3.* Of note, *EDN1, CCL2, FOS* and *JUNB* are all effectors of the TNFA signalling via NFκB pathway which is enriched in the V_Inner and V_Mid domains (Fig.28 and Fig.29) indicating overlap between blood pressure regulation and inflammatory signalling.

The gene with the most network interactions is *GNG2* (7 interactions), a G protein gamma subunit required for GTPase activity and for G protein-effector interactions (Uniprot). As G protein-coupled receptor signalling is a common mechanism of signal transmission it is perhaps unsurprising that it is a well-connected gene. Nonetheless, it may be an important signal effector in this cell layer. The chemokine *CXCL2* and chemokine receptor *CCR10* are also highly connected (6 interactions) to the same genes as one another.

There is a highly connected cluster of genes involved in mitochondrial translation comprising *MRPL46, MRPL38, MRPL24, ERAL1* and *AURKAIP1.* Translation within the mitochondria is largely for synthesis of oxidative phosphorylation complex subunits (Ott, Amunts and Brown, 2016). Indeed, oxidative phosphorylation is a ventrally enriched Mid pathway (Fig.29). As with the Inner layer there are 3 genes encoding histone proteins; *HIST1H3H, HIST1H2BG* and *HIST1H3E* also suggesting that V_Mid cells are dividing more than D_Mid cells. As cell division is a high energy process, mitochondrial biogenesis must increase which may explain the increase in mitochondrial translation.

Outer layer. In the V_Outer layer there were few interacting genes. However, the GATA zinc finger TFs are both present and linked to the BMP inhibitor Noggin which downregulates BMP4 signalling in developing HSCs (Souilhol *et al.*, 2016a).

3.4 Discussion

3.4.1 Defining optimal conditions for LCM-Seq

LCM with the aim of extracting RNA is tricky due to the high susceptibility of RNA to degradation and the multiple opportunities for exposure of the tissue to nucleases during the process. The most critical of these opportunities is the staining protocol as the tissue is submerged in solutions which may harbour nucleases and encourage their activity. A short protocol, with nuclease free, cold reagents is ideal. By trialling different staining protocols, the rapid haematoxylin and eosin stain was revealed to be the most optimal as the impact on RNA integrity could be kept to a minimum albeit at a trade-off with the enhanced cell identification using immunohistochemical methods. However, morphological information, such as the elongated endothelial cells, and staining patterns, e.g a pink layer of eosin demarcating the smooth muscle cells around the dorsal aorta, helped to define more precisely the cells taken for RNA-Seq analysis.

The efforts to optimise the LCM-Seq protocol also revealed the PicoPure RNA Isolation Kit (Thermo Fisher) as the best kit (among tested) for extracting RNA from laser captured low cell numbers for preparation of RNA-Seq libraries. The most effective method overall was taken from Nichterwitz et al, 2016 based on Smart-seq2(Picelli *et al.*, 2014) which amplifies the RNA directly from the cell lysate without prior RNA purification. This method was far more efficient than a kit purification + SMARTer® kit method enabling libraries to be prepared from as little as 20 cells rather than ~525 cells.

The measures described to optimise this technique have allowed application of a difficult technique to be applied successfully to important rare human embryo samples and to obtain meaningful spatial-transcriptomics data of the human HSC developmental niche.

3.4.2 The duplicates dilemma

An issue with generating RNA-Seq libraries from small input quantities is that the high number of PCR amplification cycles required often leads to a high number of duplicate fragments which was the case for the LCM-Seq datasets in this study. The effect of the duplicates was particularly apparent in the CS17 dataset where sample-to-sample distance clustering reflected level of duplication. Recent studies have investigated the effects of duplicates in datasets and weighed the consequences of their removal (Parekh *et al.*, 2016; Klepikova *et al.*, 2017). A critical issue is that duplicates are not totally the result of amplification bias but may result from a fragmentation bias as both enzymatic and heat fragmentation is not completely random, and natural sampling of independent molecules with the

same sequence termed 'natural duplicates' which are likely for highly expressed genes (Parekh *et al.*, 2016). Computational methods are therefore limited in defining what is a true PCR duplicate and removal has been shown to increase the false discovery rate and reduce the power in finding differentially expressed genes (Parekh *et al.*, 2016; Klepikova *et al.*, 2017).

Of note, there was always 10-30% less duplicates in the 2nd read (R2) of sequenced libraries indicating that this portion of the duplicates were due to amplification of adapter sequences in the 1st read (R1). As these will not align to the genome, they will not be included in the analysis so will not have an effect on the differential expression analysis.

3.4.3 **Spatial transcriptomics reveals signals polarised along the AGM dorsal-ventral axis**

It is well documented that the AGM is a complex niche for HSCs with many signalling pathways, cell types, ECM proteins and biochemical stimuli collaborating to provide a supportive environment for HSC generation and maintenance. Due to the interconnected nature of cells within the AGM niche, when profiling their gene expression, it is important to retain their spatial organisation. By contrast, profiling sorted cell populations loses critical region-specific gene expression and the ability to define signalling pathways contributing to the niche breaks down. As HSC emergence in the AGM is a process polarised to the ventral domain of the dorsal aorta, retaining this spatial context is particularly important. To this end, LCM-Seq was a very useful tool for generating spatial transcriptomes of the AGM.

i. Evidence of compartmentalised signalling along the D-V axis

The CS17 spatial transcriptome analysis of the dorsal aorta confirmed that expression of certain genes are polarised across a D-V split from the midline. Subdivision of the aortic vessel wall into 4 domains including additional lateral (side) subdomains revealed further localisation of gene expression levels for each subdomain indicating heterogenous functional roles of cells along the length of the D-V axis. TGF-β signalling, which has been found to be enriched in the ventral domain of the mouse dorsal aorta at the time of HSC emergence (Souilhol *et al.*, 2016a; McGarvey *et al.*, 2017), was enriched in the ventral domain of our dataset but in fact down-regulated in the ventro-lateral resolving this pathway to the most ventral cells. The involvement of TGF-β signalling in supporting HSC emergence is complex; in the bone marrow (BM) TGF-β signalling from Schwann cells and megakaryocytes maintains quiescence of HSCs which is critical for their long-term maintenance(Yamazaki *et al.*, 2011; Zhao *et al.*, 2014) and *in vitro* culture of BM CD34+ cells with TGF- β inhibits cell cycling and production of colony forming cells (Batard *et al.*, 2000; Garbe *et al.*, 1997; Sitnicka *et al.*, 1996). In the zebrafish embryo higher levels of TGF- β were correlated with impaired HSPC generation, which was partly rescued by TGF- β inhibition, suggesting that TGF- β suppresses HSC generation (Yang *et al.*, 2016). Furthermore, in mouse embryos, deletion of the TGF- β signalling effector Smad4 resulted in a dramatic increase in

the numbers of IAHCs (Lan *et al.*, 2014). However, a recent report showed that Runx1+ cells undergoing EHT express tgf-β1b and *null* mutants for this gene have impaired generation of Runx1+ cells (Zhang *et al.*, 2018). This was consistent with earlier reports that reduction in levels of the receptor Tgfβr2 reduced numbers of HSPC cells expressing *runx1* and *gata2b* (Monteiro *et al.*, 2016). It is conceivable that highly complex TGF-β/BMP signalling is subtly nuanced and may be involved in opposing regulatory roles in the AGM region. The concurrent upregulation and downregulation of TGF-β signalling observed in the ventral and ventro-lateral domains of the human dorsal aorta may reflect this.

Another striking difference between ventral and ventro-lateral domains was the ventro-lateral enrichment of 'neuroactive ligand receptor interactions' gene set suggesting that this subdomain may be particularly sensitive to secreted signals from the niche. Notably, the sympathetic ganglia are located in ventro-lateral positions to the dorsal aorta and provide signals that promote HSC emergence (Fitch *et al.*, 2012). In the mouse, IAHCs emerge abundantly in the ventro-lateral domains and it has been proposed that this may be a region of optimal HSC development(Taylor, Taoudi and Medvinsky, 2010; Souilhol *et al.*, 2016a). There is no evidence of preferential ventro-lateral IAHC or HSC development in the human embryo but these data suggest at least that niche signalling pathways may be spatially resolved across ventral and ventro-lateral domains.

In the dorsal and dorsal-lateral domains, there was enrichment of apoptosis related pathways such as 'TNFR1', 'FAS' and 'Ceramide pathways'. Controlled cell death is crucial during development for tissue remodelling suggesting more tissue remodelling in the dorsal domain. In E11.5 mice, apoptosis was shown to play a role in remodelling of the pharyngeal arch arteries in the head and neck region (Molin *et al.*, 2002). In the dorsal domain, apoptosis may be clearing the way for angiogenesis and remodelling of dorsal intersegmental vessels. Indeed, dorsal enrichment of angiogenesis was detected in the CS16 analysis.

ii. Gradients of gene expression along the D-V axis

During development, signalling gradients along the axes are a key mechanism defining a cell's spatio-temporal identity, thus co-ordinating the controlled growth of structures and specification of cell types. The archetypical example is of the dorsal-ventral polarisation of the neural tube by opposing gradients of Shh and BMPs resulting in the specification of cells along this axis to become different neuron types (BronnerFraser and Fraser, 1997). It has been demonstrated that the mouse AGM has opposing gradients of Noggin, SCF and Shh which were proposed to intersect to create optimal zone for HSC specification in the ventro-lateral dorsal aorta (Souilhol *et al.*, 2016a). It was therefore of interest to investigate gradients of signalling across the human dorsal aorta that may contribute to specification of HSCs. Of the 3 significant genes, 2 had gradual gradients across the D-V axis. *PCDH18,* which encodes protocadherin-18, was upregulated towards the dorsal domain. Overexpression of

Pcdh18a in zebrafish inhibited cell migration during development impairing the proper development of certain structures including the hatching gland (Aamar and Dawid, 2008). Cells also appeared more tightly associated. This may suggest that *PCDH18* is reduced in the ventral domain to allow for higher motility cell movements such as cell migration during EHT or for cell rolling along the vessel floor.

REN which encodes Renin was the most significant gene with a gradient of expression from ventral to dorsal. Its expression is very high in the ventral and ventro-lateral domains and drops sharply in the dorsal-lateral and dorsal domains. Renin is discussed more thoroughly in section 3.4.5.

3.4.4 Spatial transcriptomics reveals the complexity of signalling within the HSC developmental niche.

Critical factors regulating haematopoietic specification and development in the dorsal aorta can be secreted by the endothelium. However, perivascular and smooth muscle cells adjacent to the endothelium as well as deeper cell layers can also be involved as has been shown in zebrafish and mouse models (Fitch *et al.*, 2012; Souilhol *et al.*, 2016a; Tamplin *et al.*, 2015) . In order to explore both the near and longer-range signals, the region within the AGM including the dorsal aorta and surrounding mesenchyme as far as the gonad-mesonephros was microdissected into concentric layers radiating away from the dorsal aorta roughly comprising an endothelial/perivascular inner layer, a sub-aortic mesenchyme mid layer and an outer mesenchymal layer. Each of these layers was sub-divided into dorsal and ventral domains. This approach allowed gene expression profiles of different sub-domains to be contrasted in multiple paired combinations thus enabling differentially expressed genes to be determined between cells proximal and more distal to the dorsal aorta as well as polarised gene expression along the dorsal-ventral axis.

i. Non-polarised pathways in the vessel wall

In this analysis, niche signals can be broadly classified in 2 groups; non-polarised genes and pathways which provide 'background' signals that are important for the overall functioning of the niche cells and polarised genes and pathways which may contribute to the ventralisation and specification of HSC emergence. One role of the non-polarised signalling pathways would be to maintain the cells and structure of the vessel wall. For example, the IL-2 STAT5 signalling pathway is enriched in the Inner layer cells of the dorsal aorta. In adults, Interleukin-2 has been found to be sequestered to the endothelial and smooth muscle cells lining arteries by heparan sulfate proteoglycans (HSPG)(Miller *et al.*, 2012). During trans-endothelial migration in the adult organism leukocytes release the enzyme Heparanase which cleaves the HSPGs and releases IL-2 stimulating lymphocytes to proliferate. Additionally, the aortas of adult IL-2 KO mice had small, compacted smooth muscle cells resulting in much thinner vessel walls and structural break down of the endothelial cells suggesting this cytokine is required for proper vessel wall maintenance and it is tempting to speculate that it has a similar role

in the embryonic dorsal aorta. This functional role does not necessarily decouple it from supporting HSC development, however. A recent study showed that HSCs in the bone marrow of *Il2-/-* mice lose their quiescence and undergo expansion but with progeny significantly biased towards the myeloid lineage (Giampaolo *et al.*, 2017). This gives evidence for a further role of IL-2 in maintenance of haematopoiesis and regulation of differentiation. If these roles for IL-2 are found to be translatable to developmental haematopoiesis it would make IL-2 a candidate for manipulation of HSC derivation cultures to modulate expansion and skew differentiation to a preferred lineage.

Analysis of the gradated changes in gene expression towards and away from the dorsal aorta revealed a high number of DE genes encoding components of the ECM. This suggests a significant change in the ECM composition in the cells near the dorsal aorta compared with the outer mesenchyme. Cell-matrix interactions are an important part of the niche influencing stem cell properties and culture systems that more closely mimic the 3-Dimensional (3D) architecture and ECM make-up of the native niche are being developed across the stem cell field (Discher, Mooney and Zandstra, 2009). One of the genes enriched towards the vessel wall, *FN1*, which encodes fibronectin, is frequently used in cell culture. Coating 3D scaffolds with fibronectin has been shown to improve expansion of UCB HSCs (Feng *et al.*, 2006; Mousavi *et al.*, 2015). It would be interesting to investigate whether other aortic ECM proteins or proteoglycans such as *HAPLN3* (Hyaluron and Proteoglycan Link Protein 3), *BGN* (Biglycan), *ELN* (Elastin) or *COL4A1* (Collagen Type IV alpha) may improve HSC derivation or expansion cultures.

ii. Dorsal-ventral polarised signalling in the dorsal aorta vessel wall and sub-aortic mesenchyme

The data showed that there is a substantial overlap in signalling pathways between the Inner and Mid layers highlighting the importance of communication between the cells of the vessel wall and the sub-aortic mesenchyme. This overlap was particularly pronounced in the ventral domain suggesting that this region of broader signalling patterns is conducive to promoting HSC generation. Furthemore, overlapping pathways including TGF-β and TNFα signalling via NFκB are ventrally enriched in both layers. This adds evidence for a role of the sub-aortic mesenchyme in signalling to the haematogenic endothelium and it is useful to study what signals are generated here.

As the ventral domain of the dorsal aorta is the site of generation for HSCs, the dorsal domain could be expected to have reduced signalling activity in comparison. Far from this being the case, the dorsal domain has many more pathways enriched compared to the ventral domain. As previously discussed, data from Souilhol et al., shows signalling from the AoD is also important for promoting HSC generation (Souilhol *et al.*, 2016a). The D_Inner and D_Mid domains have enrichment of inflammatory signalling pathways such as cytokine and chemokine receptor interaction pathways and IL6 JAK STAT3 signalling. Many cytokines contribute to the enrichments of these pathways (data not shown). Given evidence in the literature that inflammatory pathways are involved in embryonic HSC development, the dorsal

domain may be an important source of relevant cytokines and it would be useful to investigate how these cytokines influence HSC emergence.

The GSEA provides a wealth of information regarding the patterns of signalling pathways across the dorsal aorta and surrounding mesenchymal tissue. This signalling map of the HSC developmental niche highlights not only potentially important fundamental information but also will be a valuable resource for identification of potential molecular effectors for enhancement of haematopoietic differentiation of ES cell culture systems. It would also be useful to map the transcriptomes of AGMs from earlier embryos at CS13 when IAHCs first begin to appear as this may show the signalling patterns that specify the endothelial cells to become haematogenic which may be distinct from the cues that promote HSC emergence from the haematogenic endothelium.

3.4.5 **Renin/Endothelin – Blood pressure regulators**

One gene with significant localisation to the ventral domain is *REN* which encodes Renin, a secreted protein component of the Renin-Angiotensin system (RAS). The classical physiological role of this system is of blood pressure regulation and fluid volumes balance (Sparks *et al.*, 2014). In this system, renin, an aspartyl protease, is released from the kidney in response to changes in blood pressure and fluid volume. It cleaves the N-terminus of angiotensinogen (Ang), a protein predominantly generated in the liver and found abundantly in blood serum, to convert it to angiotensin I (AngI). Angiotensin-converting-enzyme (ACE) then converts AngI to AngII by cleaving additional amino acids from the C-terminus. AngII then acts through 2 receptors AT_1 and AT_2 to stimulate vasoconstriction of blood vessels thereby modulating blood pressure or release of aldosterone from the adrenal glomerulosa triggering reabsorption of sodium into the kidneys (Sparks *et al.*, 2014)(Fig.33). The effect of RAS on blood pressure has made it an important target of pharmacological modulation for cardio-vascular therapies and ACE inhibitors have been extensively used to treat hypertension, heart failure, coronary artery disease, diabetes and chronic kidney disease (Messerli *et al.*, 2018).

In addition to regulation of blood pressure and hydro-electrolyte balance, roles for RAS in regulation of angiogenesis and inflammation have also been documented (Duprez, 2006). Importantly, there is an emerging role of RAS in the haematopoietic niche. In the bone marrow, AngII stimulates release of arachidonic acid which mediates haematopoietic progenitor apoptosis and proliferation (Richmond *et al.*, 2004). ACE-KO mice revealed that ACE regulates myeloid and erythroid proliferation and maturation via AngII (Lin *et al.*, 2011) and, consistently, lack of ACE results in anemia in the adult mice (Lin *et al.*, 2011; Cole *et al.*, 2000). Accordingly, ACE was found to be expressed in the chick yolk sac endoderm and pharmocological inhibition disrupted primitive erythropoiesis by reducing erythroblast proliferation (Savary *et al.*, 2005).

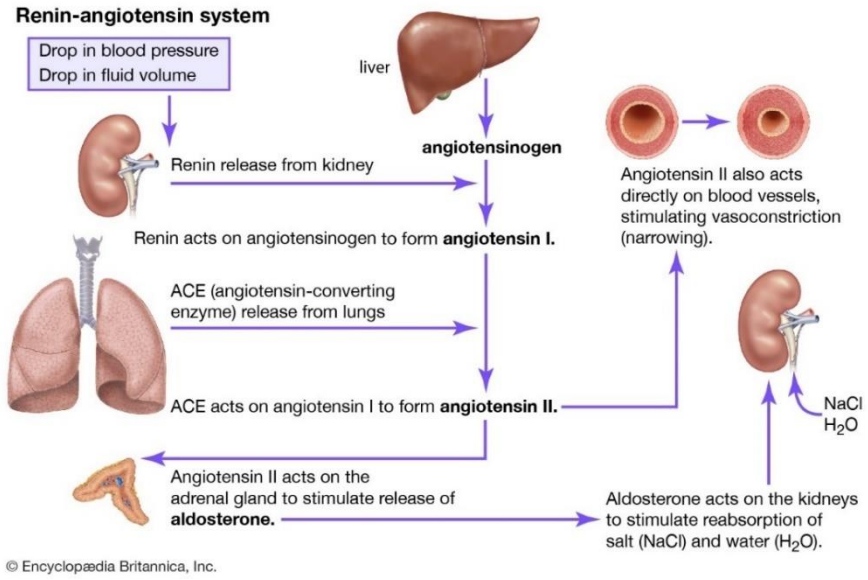


Figure 33. The classical Renin-Angiotensin system (Encylopaedia Britannica Inc.). Renin is released from the kidney in response to a detected drop in blood pressure or fluid volume. It acts on angiotensinogen to convert it to angiotensin I which is in turn converted to angiotensin II by the enzyme angiotensin-converting enzyme (ACE). Angiotensin II stimulates vasoconstriction of the blood vessels or stimulates release of aldosterone from the adrenal glands. Aldosterone acts on the kidneys to stimulates reabsorption of NaCl and H_2O.

As discussed in the introduction, ACE has been reported to mark migrating cells from the P-Sp (the presumptive haematogenic endothelium) which colonise the ventral endothelium of the dorsal aorta and specify or directly contribute to emerging HSCs (Jokubaitis *et al.*, 2008). The role of ACE in supporting HSC development was further investigated using a human embryoid body (hEB) haematopoeitic differentiation culture system (Zambidis *et al.*, 2008). Blast colony-forming cell (BL-CFC) colonies were derived only from the ACE+ fraction of hEBs and, interestingly, inhibition of both ACE and AGTR2 significantly reduced the number of colony forming units (CFUs) and replated blast colonies produced predominantly endothelial cells. Conversely, AGTR1 inhibition greatly enhanced CFU production and replated cells gave rise to haematopoeitic progeny. This suggests opposing receptor signalling is critical for determining an endothelial vs. haematopoietic fate to haematogenic endothelial cells.

Taken together, these data suggest a role of RAS in promoting HSC emergence in the AGM region. However, thus far only ACE and downstream effectors have been investigated and not Renin itself. Notably, *REN* and the receptor *AGTR2* were most highly expressed in the ventral sub-aortic mesenchymal layer. This position, not directly adjacent to the endothelial ACE+ cells but deeper into the underlying cells, suggests Renin may not act exclusively on the ACE+ clusters adhering to the endothelium. Indeed, Renin-prorenin receptor mediated signalling has been shown to have roles independent of AgtII generation stimulating TGF-β signalling and subsequent increases in matrix

proteins fibronectin, collagen-1 and plasminogen activator inhibitor-type 1 (Huang *et al.*, 2006). Expression of the prorenin receptor (ATP6AP2) is seen throughout the region of the AGM mapped in this analysis (data not shown).

The gene *EDN1* encoding another blood pressure regulating peptide endothelin-1 (Edn1) increased in expression towards the dorsal aorta and was also significantly enriched in the ventral sub-aortic mesenchymal cells. Endothelin-1 (ET-1) is a potent vasoconstrictor which acts through the G-protein coupled transmembrane receptors ET_A and ET_B (Haque, Welch and Loizidou, 2013). There are significant overlaps in the RAS and Edn1 pathways; both AngII and Edn1 stimulate arachidonic acid release and subsequent prostaglandin synthesis (Richmond *et al.*, 2004; Yousufzai and AbdelLatif, 1997), and AngII has been reported to upregulate TNF-α synthesis (Lin *et al.*, 2011; Yamada *et al.*, 2010) which is known to stimulate ET-1 production (Haque, Welch and Loizidou, 2013). Indeed, there have been several reports of induction of Edn1 expression and protein production by adding AngII to endothelial (Chua *et al.*, 1993; Imai *et al.*, 1992) and vascular smooth muscle cell cultures (Hahn *et al.*, 1990; Sung *et al.*, 1994). Conversely, Edn1 has been shown to inhibit Renin production in the main Renin producing cells, the juxtaglomerular cells, in the kidney (Ackermann *et al.*, 1995; Beierwaltes and Carretero, 1992; Lin *et al.*, 1993). Endothelin-1 can also stimulate aldosterone production as potently as AngII (Rossi *et al.*, 1997). There is clearly crosstalk between the two pathways and the ventral enrichment of both Renin and Endothelin-1 makes a localised and co-ordinated role for the two in the AGM an intriguing possibility.

Interestingly, Edn1 is ventrally localised in the pharyngeal arches and, along with BMP, is required for conferring cells their positional identiy along the dorsal-ventral axis and enabling development of the ventral craniofacial structures (Alexander *et al.*, 2011). It is interesting that it is also ventrally localised in relation to the dorsal aorta, as is BMP, and a ventralising effect of Edn1 in the dorsal aorta would be worth investigating. The spatial polarisation of both Renin and Edn1 along with the implications of the RAS in haematopoiesis and the involvement of both factors in ventrally enriched pathways (TGFβ and TNFa signalling via NFκB) make them promising candidates as players in the HSC developmental niche.

4. Resolution of the spatial transcriptome to defined cell populations

4.1 Transcriptome analysis of spatially resolved, sorted haemato-endothelial populations from the human dorsal aorta.

The spatial transcriptome dataset provided a map of signalling across the aortic HSC niche. In order to understand how these niche signals may interact with the haematogenic endothelium to influence HSC emergence, it is necessary to define the molecular signature of the cells along the EHT axis. Although LCM is a useful tool to interrogate the transcriptome of cells without losing their spatial identity, the phenotype of the cells taken is not certain. For example, the Inner layer of the CS16 LCM-Seq (Fig.21) contains both endothelial and perivascular cells and we do not know which of the genes expressed derive from which cell type. To resolve these signals to cell populations of the haematogenic endothelium the dorsal aortae from 2 embryos (N1 = CS15, N2 = CS16) were sub-dissected manually into dorsal and ventral portions (indicated by D and V) and sorted based on VC and CD45 expression giving VC+CD45- endothelial cells, VC+CD45+ Haematopoietic Stem/Progenitor Cells (HSPCs), and VC-CD45+ haematopoietic cells (Fig.34). From these cells, the bulk RNA was extracted and RNA-Seq libraries were prepared and sequenced and DGE and GSEA analysis was carried out. This dataset not only allows the dynamics of gene expression during EHT to be elucidated but also to resolve signalling pathways from the LCM-Seq analysis such as the renin and endothelin pathways.

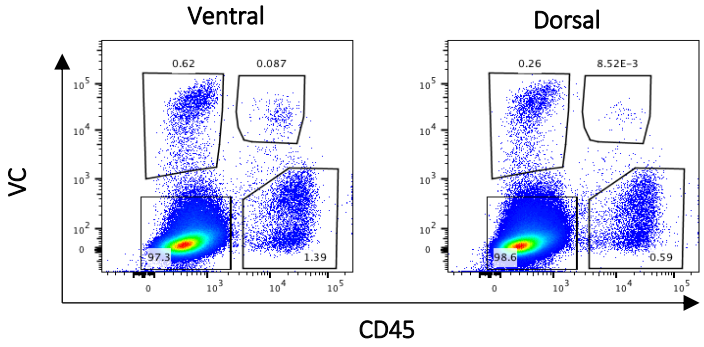


Figure 34. Sorting strategy for transcriptome analysis of haemato-endothelial populations from the dorsal aorta. The dorsal aortae from CS15/16 embryos (n=2) were sub-dissected into ventral and dorsal domains and sorted based on the expression of VE-Cadherin (VC) and CD45 to give (A) VC+CD45- endothelial, (B) VC+CD45+ haematopoietic stem/progenitor population (HSPC) and (C) VC-CD45+ haematopoietic population for subsequent RNA-Sequencing.

4.1.1 Visualisation and quality assessment of the sequenced read data by sample clustering analysis

The low cell numbers and subsequent high number of PCR amplification cycles resulted in high duplicate levels (Table 8). Duplicate levels are the highest for the D_VC+CD45+ populations which also had the lowest number of cells. High duplicate levels are characteristic for low input RNA due to the

low starting complexity of the amplification reactions (Parekh *et al.*, 2016). On the whole, duplicate levels do not seem to greatly affect the molecular signatures of the cell populations as different cell types from 2 embryos cluster together on the PCA analysis with the exception of D_VC+CD45+ which differ greatly from each other and the other cell populations along PC1 and PC2 (Fig.35a). Along the PC3 axis they split to cluster most closely with either the VC+CD45- or VC-CD45+. This indicates that the cells in the 2 D_VC+CD45+ populations are not at all similar and have been excluded from DGE analysis. The heatmap of sample-to-sample distances and of the top 500 genes by expression level also indicates the lack of correlation between these samples (Fig.35b, c). The V_VC+CD45+ HSPC population clusters more closely to the VC-CD45+ haematopoietic populations along PC2 but falls directly between the VC+CD45- and VC-CD45+ populations along PC3 demonstrating its intermediate position in EHT (Fig.35a). There is no strong indication of D-V differences on the PCA analysis except for between V_VC+CD45+ and D_VC+CD45+ populations and, as mentioned, this may be a duplication level artefact.

Specimen	Sample_Name	Input cell numbers	% Dups
	Ventral VC+CD45-_1	2921	88.10%
	Ventral VC+CD45-_2	2921	84.10%
	Ventral VC+CD45+_1	374	91.80%
	Ventral VC+CD45+_2	374	87.80%
	Ventral VC-CD45+_1	5056	82.50%
	Ventral VC-CD45+_2	5056	77.80%
	Dorsal VC+CD45-_1	2661	82.40%
	Dorsal VC+CD45-_1	2661	77.70%
	Dorsal VC+CD45+_1	86	94.40%
	Dorsal VC+CD45+_2	86	91.50%
	Dorsal VC-CD45+_1	4558	84.80%
1	Dorsal VC-CD45+_2	4558	79.20%
	Ventral VC+CD45-_1	1199	84.30%
	Ventral VC+CD45-_2	1199	78.80%
	Ventral VC+CD45+_1	274	88.30%
	Ventral VC+CD45+_2	274	80.80%
	Ventral VC-CD45+_1	4120	75.80%
	Ventral VC-CD45+_2	4120	68.90%
	Dorsal VC+CD45-_1	183	87.90%
	Dorsal VC+CD45-_1	183	82.60%
	Dorsal VC+CD45+_1	8	92.30%
	Dorsal VC+CD45+_2	8	89.40%
	Dorsal VC-CD45+_1	1058	77.50%
2	Dorsal VC-CD45+_2	1058	69.20%

Table 8 Table of sorted cell populations for each sample taken for RNA-Sequencing with correlating input cell number and duplicate levels following sequencing. For each sample read 1 = _1 and read 2 = _2

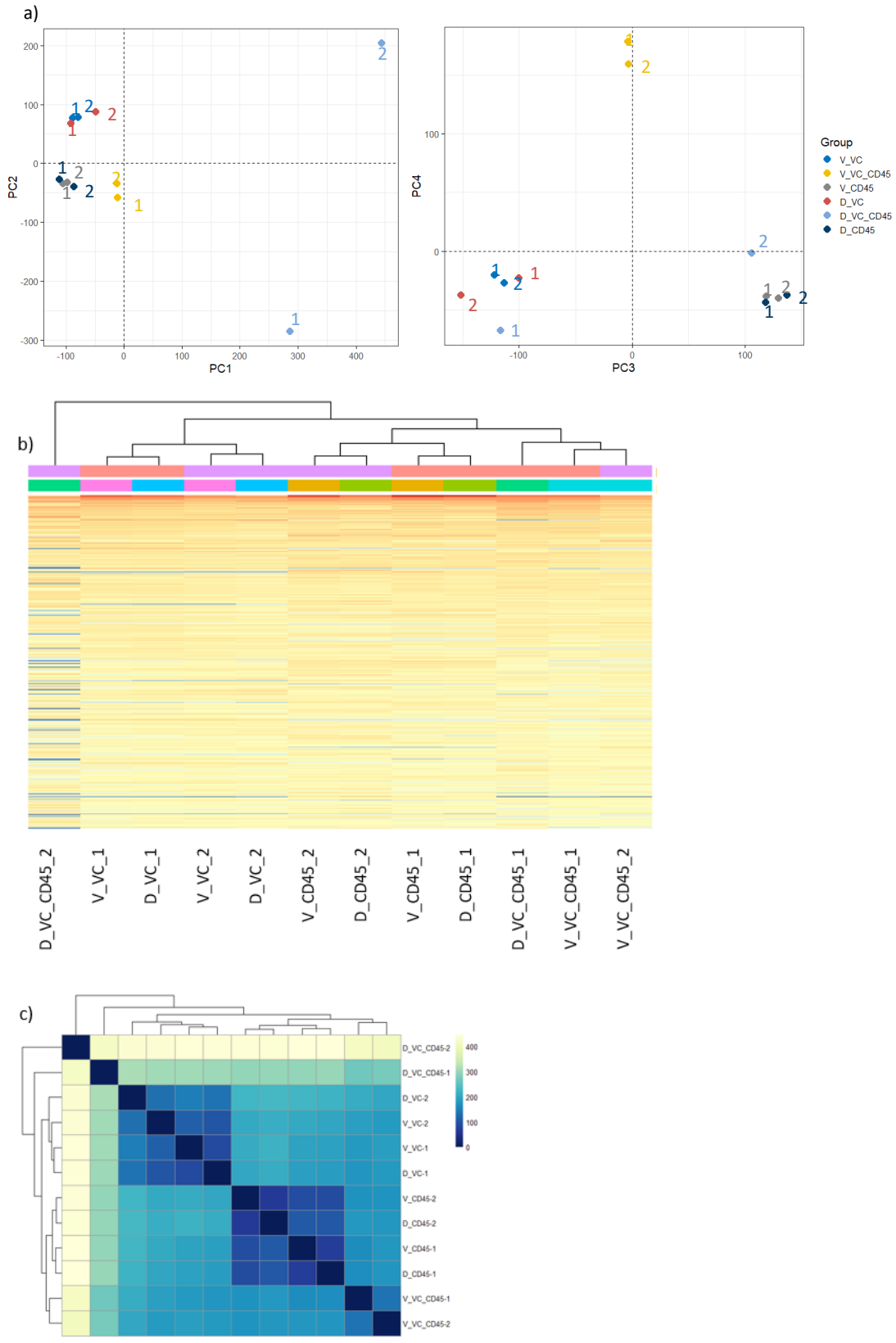


Figure 35. Clustering visualisation of dorsal aorta haemato-endothelial population transcriptomes. V = ventral. D = Dorsal. VC = VC+CD45-, VC_CD45 = VC+CD45+, CD45 = VC-CD45+ a) Principle component analysis of sequenced cell populations, principle components (PC) 1 and 2 (left) and PC3 and 4 (right). b) Heatmap of top 500 genes by mean expression level clustered by sample similarity. c) Heatmap of sample-to-sample distances

4.1.2 **Dorsal-ventral differences are evident in the endothelial VC+CD45- population, including ventrally enriched EDN1.**

There were 94 DE genes between the V_ VC+CD45- and D_VC+CD45- populations. This included *EDN1* which was also ventrally enriched in the Mid layer of the LCM-Seq spatial transcriptome analysis (Fig.36, red box and Fig. 32 from chapter 3). It was not significantly enriched in the ventral domain of the Inner layer however (log2FC = 2.4, p.adj = 0.4) which is where most endothelial cells are found. This indicates that more precise sorted populations were required to find these D-V differences in the endothelial population which may have been occluded by DE genes in the perivascular cells. Indeed, there is only 1 overlapping gene between the ventral vs dorsal VC+CD45- and Inner layer LCM-Seq datasets which is *PAX1* – a dorsally enriched gene in both datasets (Fig.36, green box). This is a TF which, in the mouse, is expressed in the sclerotome between E9.5 and E11.5 in the mesenchyme dorsally adjacent to the dorsal aorta in response to Shh signalling from the notochord (Furumoto *et al.*, 1999). It is reasonable then that it should also be expressed in the dorsal cells of the aorta. Other TFs were defined using the GO term 'DNA-binding transcription factor activity' GO:0003700 and are indicated with an asterisk (Fig.36) and include ventrally enriched *GATA4* and *TBX5*.

A proportion of the genes enriched in the D_VC+CD45- are haemoglobin genes (HB*) which might indicate that these populations were contaminated with erythrocytes which have highly abundant globin mRNA. However, bona fide expression of these genes in the D_VC+CD45- populations should not be ruled out. Notably the haemoglobins are embryonic (*HBE1*, *HBZ*) and foetal (*HBG1, HBG2)* sub-types.

4.1.3 **Dynamic expression of genes across the endothelial-to-haematopoietic transition**

By finding differentially expressed genes between the endothelial, HSPC and haematopoietic cell populations in the ventral domain, the gene expression changes during EHT could be elucidated including TFs, secreted factors, and cell receptors (Fig.37a-c). In total the ventral HSPC population has 1050 and 910 genes upregulated compared to the endothelial and haematopoietic populations respectively (Fig.37d). Out of these genes, 251 are overlapping indicating that they are upregulated in the transition from VC+CD45- to VC+CD45+ then downregulated to VC-CD45+ providing an HSPC specific 'signature'. Some of the genes within this signature can be clearly visualised in the heatmaps (Fig.37a-c). For example, the TFs *MYB* and *SETBP1* are strongly upregulated only in the HSPC population whereas expression of others such as *RUNX1*, *IKZF1*, and *SPI1* continue to be expressed in the more mature haematopoietic cells. There are also a number of genes which are downregulated on the HSPC population compared to both endothelial and haematopoietic populations such as the TFs *ZFP36L1* and *BHLHE40* and the secreted factors *NRP1* and *A2M* suggesting inhibition of these factors is required for HSPC production from endothelial cells but not for their maturation. The HSPC population already

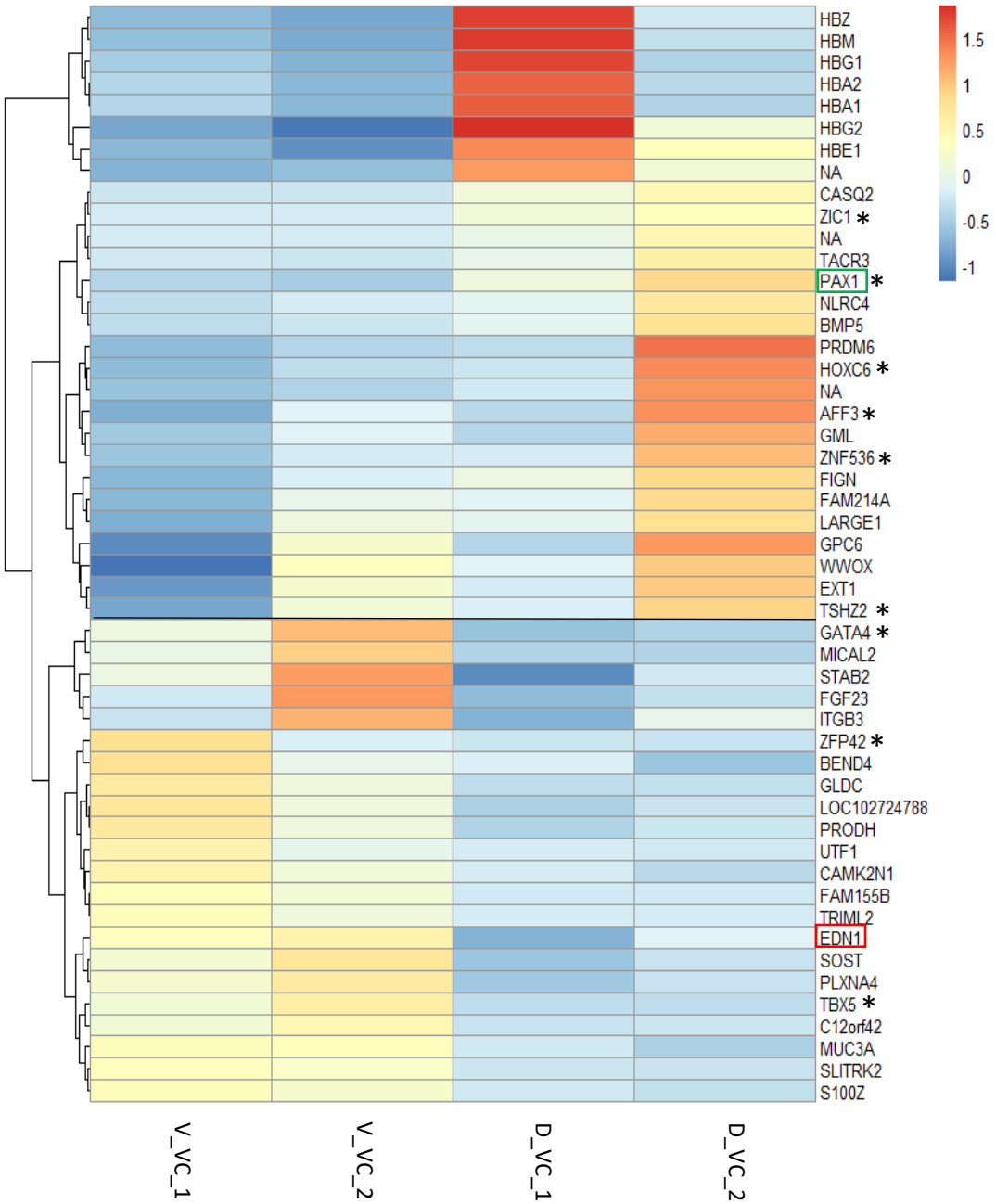


Figure 36. Dorsal-ventral polarised gene expression in the dorsal aorta endothelial cells. V = ventral. D = Dorsal. VC = VC+CD45-. Heatmap of top 50 differentially expressed genes between V_VC and D_VC populations p.adj = <0.05. Green box highlight *PAX1,* an overlapping gene with LCM-Seq Inner analysis. Red box indicates *EDN1* a gene ventrally enriched in the LCM-Seq Mid analysis. * = transcription factors. NA = ensembl IDs without gene identifiers.

expresses haematopoietic markers such as *CD84*, *CSF3R*, and *CSF1R* indicating that some cells within this population are lineage primed.

To statistically define gradated monotonic change in gene expression across EHT (VC+CD45- → VC+CD45+ → VC-CD45+) a likelihood ratio test (LRT) was used instead of the default Wald test of DESeq2. This determined 3615 significantly differentially expressed genes along the EHT axis (p.adj<0.01) some of which are shown in Fig.38. As to be expected *CDH5* (VC) decreases in expression from VC+CD45- to VC-CD45+ and *PTPRC* (CD45) increases. For some genes such as *ADGRF5* and *G0S2* the change in expression across the EHT axis is gradual whereas others decrease or increase sharply from one population to the next. For example, *ADGRL4* and *PRND* drop steeply in expression level between the endothelial and HSPC populations and stay much the same between the VC+CD45+ and VC-CD45+ populations. These steeper changes in gene expression may indicate a switching on or off of genes required for HSPC specification from the endothelium whereas more gradual changes could indicate secondary changes.

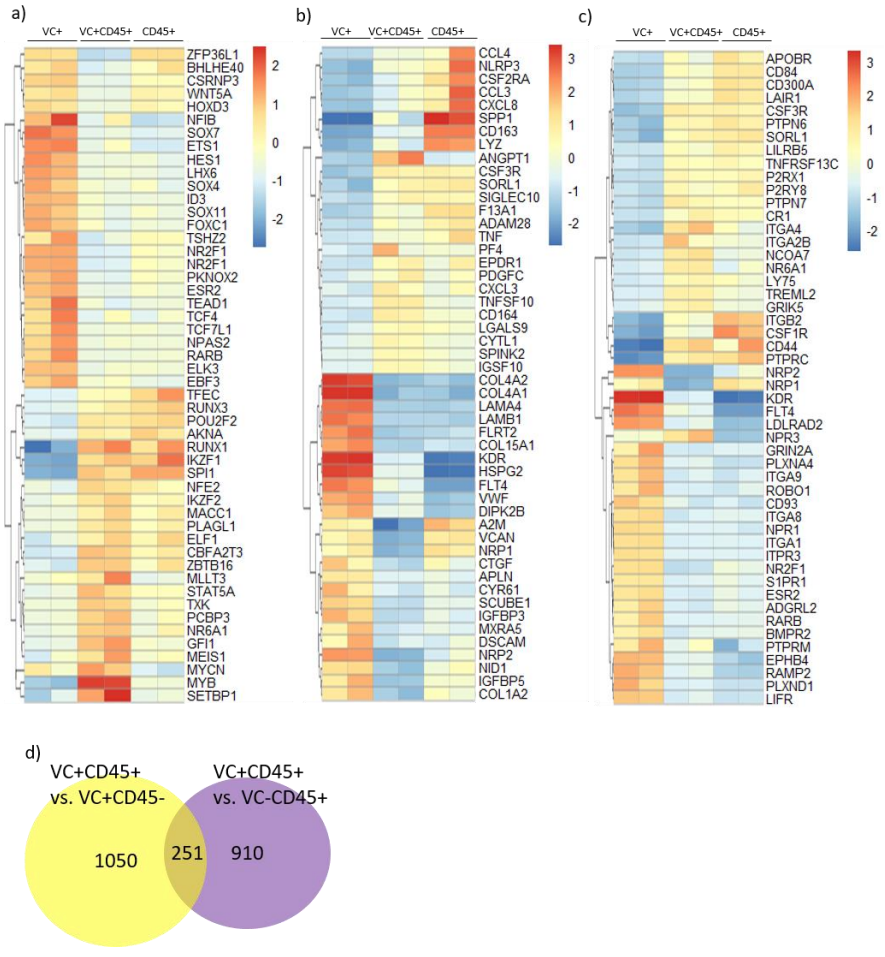


Figure 37. Differentially expressed genes across the EHT axis. VC+ = VC+CD45-. CD45+ = VC-CD45+ Heatmaps of top 25 enriched genes for V_VC+CD45- vs. V_VC+CD45+ and vice versa for a) Transcription Factors, b) Secreted Factors, c) Cell receptors. p.adj = <0.05. d) Venn diagram of genes enriched in the ventral HSPC VC+CD45+ population vs. VC+CD45- and VC-CD45+

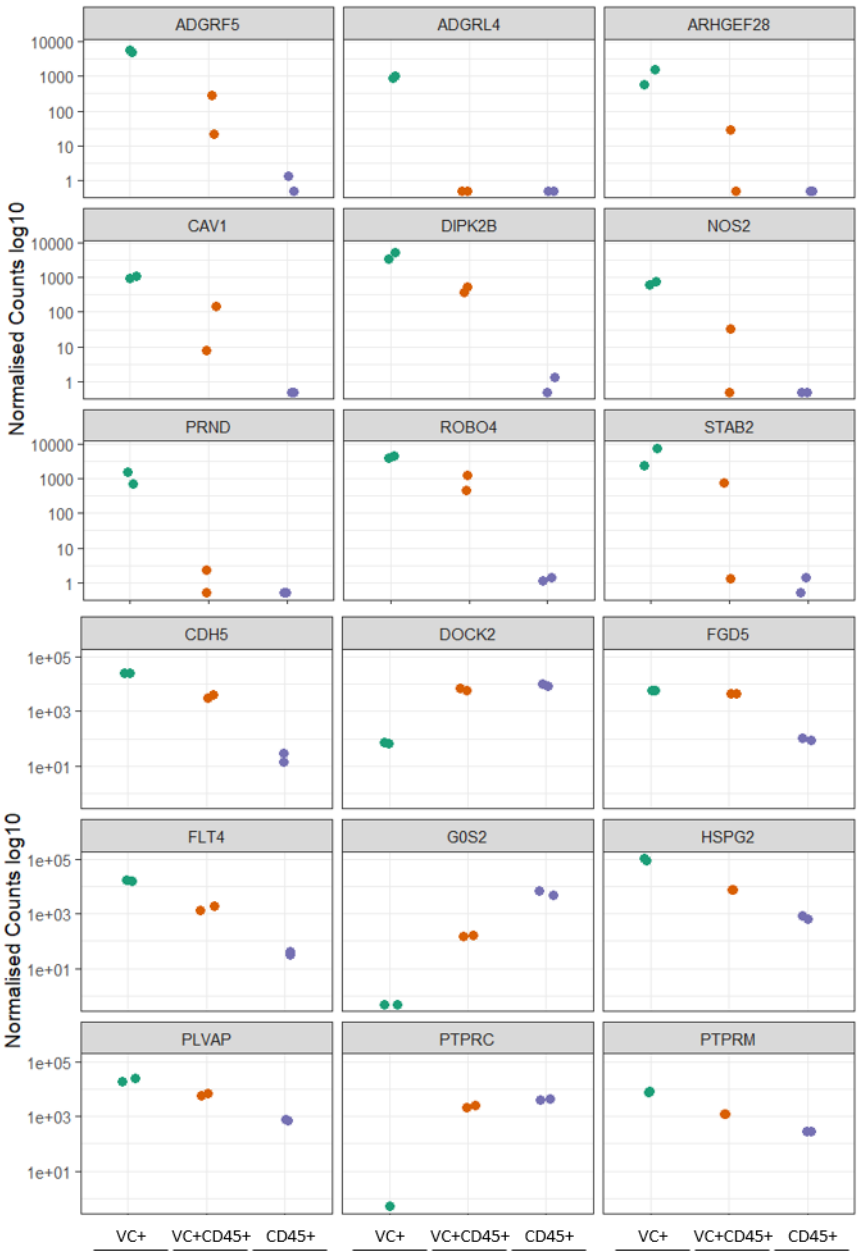


Figure 38. Gradients of gene expression across the EHT axis. Graphs of normalised counts for the top 24 differentially expressed genes determined using a likelihood ratio test to define gradated changes across the 3 populations. VC+ = VC+CD45-, VC+CD45+ = VC+CD45+, CD45+ = VC-CD45+. P.adj<0.05

4.1.4 Gene set enrichment analysis reveals pathways differentially enriched along EHT axis

GSEA of the differentially expressed genes between sub-populations in the ventral domain can reveal pathways involved in EHT and integration of these pathways with the GSEA from the LCM-Seq (Chapter 3.3.5) may indicate how signalling pathways from the niche interact with these populations to promote HSC emergence.

Gene sets enriched in the endothelial population compared with HSPC include more structural and extracellular matrix related gene sets such as 'Core Matrisome' (NABA), 'Extracellular Matrix Organisation', 'Collagen Formation' (Reactome), 'Collagens' (NABA), and 'Basement Membranes' (NABA) which is to be expected from the structural integrity of the endothelium and its interactions with the underlying cells of the vessel wall (Fig.39a). It follows that in the HSPC population gene sets related to haematopoietic cell function are enriched such as 'Hematopoietic Cell Lineage' (KEGG), 'Immune System' (Reactome), 'Immunoregulatory Interactions Between a Lymphoid and a Non Lymphoid Cell' (REACTOME) to name just a few (Fig.39b).

In the LCM-Seq analysis 'TGF Beta Signalling' (Hallmark) and 'TNFA signalling via NFKB' (Hallmark) were ventrally enriched in both the inner endothelial/perivascular layer and the mid sub-aortic mesenchyme (Fig.28,29). In this dataset 'TGF Beta Signalling' (Hallmark) is enriched on the endothelial population and 'TNFA signalling via NFKB' (Hallmark) is enriched on the HSPC population indicating that these niche pathways may influence distinct cell populations (Fig.39a, b). Multiple pathways enriched in cells of the vessel wall as determined by the LCM-Seq Inner vs Outer analysis are also differentially enriched on the endothelial and HSPC populations. 'Notch Signalling' (Hallmark) is enriched on the endothelial population whereas 'IL6 JAK STAT Signalling' is HSPC enriched. Of note, 'Hedgehog signalling' is also enriched on the endothelial population. This is consistent with reports that hedgehog signalling is required for dorsal aorta development and arterial specification (Gering and Patient, 2005).

There are much fewer pathways enriched in the HSPC population when compared with the VC-CD45+ haematopoietic population (Fig.40a). This may indicate that many of the enriched HSPC pathways vs. the endothelial population derive from the acquisition of mature haematopoietic cell traits rather than the underlying specification process. Several of the pathways that are enriched in the HSPC population compared to both endothelial and haematopoetic populations (Fig.40a, indicated with an asterisk) are involved in translation e.g. 'Peptide Chain Elongation' (Reactome), 'Translation' (Reactome) etc. This indicates a push for high protein synthesis. 'Myc Targets V2' is also enriched on the HSPC population indicating proliferation and, along with increased protein synthesis, higher metabolism of these cells.

Interestingly, a high proportion of pathways enriched on the VC-CD45+ haematopoietic population are also dorsally enriched in Inner and Mid layers from the LCM-Seq datasets (Fig.40b, and Figs.28,29). This includes 'Chemokine Receptors Bind Chemokines' (Reactome), 'IL12 STAT4 Pathway' (PID) involved in regulation of differentiation and proliferation of T cells (Watford *et al.*, 2004) and 'CTLA4 pathway' (Biocarta) involved in modulation of T cell function (Buchbinder and Desai, 2016). This raises the question of whether signals from the dorsal domain may be supporting the maturation of certain haematopoietic cell types, perhaps through chemokine signalling. Of note, there are some pathways enriched in both the endothelial and haematopoietic populations compared with HSPC population

such as 'NOTCH signalling' (HALLMARK) and 'Hedgehog Signalling' (Hallmark and KEGG). This suggests that these pathways require downregulation during HSPC generation Notably, downregulation of Notch in mouse HSC maturation has been shown(Souilhol *et al.*, 2016b; Lizama *et al.*, 2015; Tang *et al.*, 2013).

a) **VC+CD45- (Endothelial) enriched pathways vs VC+CD45 (HSPC)**

Pathway	NES	FDR
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	1.83	0.00
HALLMARK_MYOGENESIS	1.72	0.00
NABA_CORE_MATRISOME	2.02	0.01
HALLMARK_HEDGEHOG_SIGNALING	1.55	0.01
HALLMARK_UV_RESPONSE_DN	1.50	0.01
HALLMARK_ANGIOGENESIS	1.50	0.01
HALLMARK_APICAL_SURFACE	1.55	0.02
HALLMARK_NOTCH_SIGNALING	1.51	0.02
HALLMARK_APICAL_JUNCTION	1.51	0.02
HALLMARK_KRAS_SIGNALING_DN	1.41	0.04
HALLMARK_PANCREAS_BETA_CELLS	1.41	0.04
NABA_ECM_GLYCOPROTEINS	1.81	0.05
KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	1.77	0.05
NABA_COLLAGENS	1.79	0.05
PID_SYNDECAN_1_PATHWAY	1.81	0.07
HALLMARK_GLYCOLYSIS	1.34	0.09
REACTOME_NCAM1_INTERACTIONS	1.69	0.12
NABA_BASEMENT_MEMBRANES	1.62	0.13
REACTOME_DIABETES_PATHWAYS	1.62	0.13
REACTOME_NCAM_SIGNALING_FOR_NEURITE_OUT_GROWTH	1.63	0.14
REACTOME_COLLAGEN_FORMATION	1.63	0.15
HALLMARK_TGF_BETA_SIGNALING	1.28	0.16
NABA_MATRISOME	1.63	0.16
HALLMARK_HYPOXIA	1.26	0.17
HALLMARK_COAGULATION	1.26	0.17
REACTOME_TRANSMISSION_ACROSS_CHEMICAL_SYNAPSES	1.64	0.17
REACTOME_NEUROTRANSMITTER_RECEPTOR_BINDING_AND_DOWNSTREAM_TRANSMISSION_IN_THE_POSTSYNAPTIC_CELL	1.57	0.18
REACTOME_CLASS_B_2_SECRETIN_FAMILY_RECEPTORS	1.55	0.18
REACTOME_NEURONAL_SYSTEM	1.55	0.20
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION	1.52	0.21
PID_INTEGRIN1_PATHWAY	1.53	0.21
PID_AVB3_INTEGRIN_PATHWAY	1.52	0.21
REACTOME_GPCR_LIGAND_BINDING	1.49	0.25

KEY:
Pathways overlapping with LCM-Seq
Inner vs Outer (Both Ventral and Dorsal)
Inner Ventral vs Inner Dorsal
Mid Ventral vs Mid Dorsal
Inner Dorsal vs Inner Ventral
Mid Dorsal vs Mid Ventral

Figure 39 Gene Set Enrichment Analysis for the V_VC+CD45- vs V_VC+CD45+. GSEA input was differentially expressed genes ranked by log2 fold change. a) V_VC+CD45- enriched pathways. b) V_VC+CD45+ enriched pathways (next page). Colour coding is described in the key**.** FDR<0.25.

b)	VC+CD45+ (HSPC)	
Pathway	NES	FDR
HALLMARK_ALLOGRAFT_REJECTION	-2.01	0.00
KEGG_RIBOSOME	-2.37	0.00
REACTOME_PEPTIDE_CHAIN_ELONGATION	-2.41	0.00
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	-2.42	0.00
KEGG_HEMATOPOIETIC_CELL_LINEAGE	-2.47	0.00
REACTOME_IMMUNE_SYSTEM	-2.47	0.00
REACTOME_NONSENSE_MEDIATED_DECAY_ENHANCED_BY_THE_EXON_JUNCTION_COMPLEX	-2.48	0.00
KEGG_B_CELL_RECEPTOR_SIGNALING_PATHWAY	-2.48	0.00
REACTOME_IMMUNOREGULATORY_INTERACTIONS_BETWEEN_A_LYMPHOID_AND_A_NON_LYMPHOID_CELL	-2.31	0.00
REACTOME_INFLUENZA_LIFE_CYCLE	-2.52	0.00
REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM	-2.54	0.00
REACTOME_3_UTR_MEDIATED_TRANSLATIONAL_REGULATION	-2.55	0.00
PID_BCR_5PATHWAY	-2.25	0.00
KEGG_FC_EPSILON_RI_SIGNALING_PATHWAY	-2.26	0.00
REACTOME_TRANSLATION	-2.55	0.00
REACTOME_METABOLISM_OF_RNA	-2.24	0.00
REACTOME_INFLUENZA_VIRAL_RNA_TRANSCRIPTION_AND_REPLICATION	-2.56	0.00
KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	-2.24	0.00
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	-2.21	0.00
REACTOME_SIGNALING_BY_ILS	-2.58	0.00
REACTOME_ADAPTIVE_IMMUNE_SYSTEM	-2.14	0.00
REACTOME_METABOLISM_OF_MRNA	-2.14	0.00
KEGG_LEISHMANIA_INFECTION	-2.58	0.00
REACTOME_ACTIVATION_OF_THE_MRNA_UPON_BINDING_OF_THE_CAP_BINDING_COMPLEX_AND_EIFS_AND_SUBSEQUENT_BINDING_TO_43S	-2.10	0.00
KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS	-2.10	0.00
KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	-2.08	0.00
REACTOME_IL_3_5_AND_GM_CSF_SIGNALING	-2.08	0.00
KEGG_OXIDATIVE_PHOSPHORYLATION	-2.03	0.00
HALLMARK_INTERFERON_ALPHA_RESPONSE	-1.84	0.01
HALLMARK_MYC_TARGETS_V2	-1.71	0.01
KEGG_APOPTOSIS	-1.94	0.01
PID_TCR_PATHWAY	-1.93	0.01
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS_	-1.92	0.01
PID_IL8_CXCR2_PATHWAY	-1.91	0.01
HALLMARK_IL6_JAK_STAT3_SIGNALING	-1.72	0.01
PID_IL12_2PATHWAY	-1.89	0.01
HALLMARK_INTERFERON_GAMMA_RESPONSE	-1.67	0.02
KEGG_CHEMOKINE_SIGNALING_PATHWAY	-1.83	0.02
PID_FCER1_PATHWAY	-1.83	0.02
REACTOME_INNATE_IMMUNE_SYSTEM	-1.82	0.02
REACTOME_ANTIGEN_PROCESSING_CROSS_PRESENTATION	-1.84	0.02
KEGG_T_CELL_RECEPTOR_SIGNALING_PATHWAY	-1.77	0.02
REACTOME_HOST_INTERACTIONS_OF_HIV_FACTORS	-1.77	0.02
BIOCARTA_HIVNEF_PATHWAY	-1.76	0.03
KEGG_ACUTE_MYELOID_LEUKEMIA	-1.73	0.03
REACTOME_SIGNALING_BY_THE_B_CELL_RECEPTOR_BCR	-1.71	0.03
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	-1.70	0.04
SIG_BCR_SIGNALING_PATHWAY	-1.69	0.04
REACTOME_NUCLEOTIDE_BINDING_DOMAIN_LEUCINE_RICH_REPEAT_CONTAINING_RECEPTOR_NLR_SIGNALING_PATHWAYS	-1.68	0.04
HALLMARK_INFLAMMATORY_RESPONSE	-1.44	0.04
HALLMARK_TNFA_SIGNALING_VIA_NFKB	-1.39	0.05
PID_CMYB_PATHWAY	-1.58	0.07
REACTOME_ANTIGEN_ACTIVATES_B_CELL_RECEPTOR_LEADING_TO_GENERATION_OF_SECOND_MESSENGERS	-1.55	0.09
REACTOME_PLATELET_AGGREGATION_PLUG_FORMATION	-1.55	0.09
REACTOME_CELL_SURFACE_INTERACTIONS_AT_THE_VASCULAR_WALL	-1.52	0.10
KEGG_VEGF_SIGNALING_PATHWAY	-1.50	0.11
KEGG_LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION	-1.50	0.11
KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	-1.49	0.11
REACTOME_INTERFERON_SIGNALING	-1.49	0.11
REACTOME_SIGNALING_BY_RHO_GTPASES	-1.48	0.11
REACTOME_GPVI_MEDIATED_ACTIVATION_CASCADE	-1.46	0.13
PID_CD8_TCR_PATHWAY	-1.44	0.13
PID_CXCR4_PATHWAY	-1.44	0.14
REACTOME_HEMOSTASIS	-1.44	0.14
KEGG_PARKINSONS_DISEASE	-1.42	0.15
KEGG_JAK_STAT_SIGNALING_PATHWAY	-1.41	0.15
PID_IL8_CXCR1_PATHWAY	-1.41	0.15
REACTOME_HIV_INFECTION	-1.39	0.16
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	-1.38	0.17
HALLMARK_IL2_STAT5_SIGNALING	-1.24	0.19
KEGG_CELL_ADHESION_MOLECULES_CAMS	-1.35	0.19
KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM	-1.35	0.19
REACTOME_PLATELET_ACTIVATION_SIGNALING_AND_AGGREGATION	-1.34	0.19
REACTOME_NRAGE_SIGNALS_DEATH_THROUGH_JNK	-1.33	0.20
PID_IL4_2PATHWAY	-1.34	0.20
PID_PI3KCI_PATHWAY	-1.33	0.20
PID_AMB2_NEUTROPHILS_PATHWAY	-1.30	0.23
REACTOME_CLASS_I_MHC_MEDIATED_ANTIGEN_PROCESSING_PRESENTATION	-1.29	0.24

a)

	VC+CD45+ (HSPC) vs VC-CD45+ (Haematopoietic)		
	Pathway	NES	FDR
*	**REACTOME_PEPTIDE_CHAIN_ELONGATION**	2.30	0.00
*	REACTOME_3_UTR_MEDIATED_TRANSLATIONAL_REGULATION	2.20	0.00
*	**KEGG_RIBOSOME**	2.08	0.02
	REACTOME_FORMATION_OF_THE_TERNARY_COMPLEX_AND_SUBSEQUENTLY_THE_43S_COMPLEX	2.08	0.02
*	REACTOME_NONSENSE_MEDIATED_DECAY_ENHANCED_BY_THE_EXON_JUNCTION_COMPLEX	2.03	0.02
*	REACTOME_ACTIVATION_OF_THE_MRNA_UPON_BINDING_OF_THE_CAP_BINDING_COMPLEX_AND_EIFS_AND_SUBSEQUENT_BINDING_TO_43S	2.01	0.02
*	HALLMARK_MYC_TARGETS_V2	1.54	0.03
	REACTOME_INTEGRIN_ALPHAIIB_BETA3_SIGNALING	1.88	0.06
*	REACTOME_INFLUENZA_VIRAL_RNA_TRANSCRIPTION_AND_REPLICATION	1.89	0.07
*	REACTOME_TRANSLATION	1.80	0.10
*	REACTOME_PLATELET_AGGREGATION_PLUG_FORMATION	1.74	0.16
	PID_FANCONI_PATHWAY	1.70	0.18

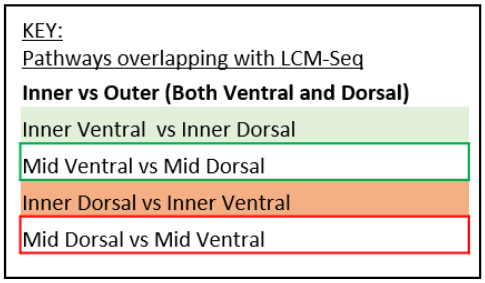


Figure 40 Gene Set Enrichment Analysis for the V_VC+CD45+ vs V_VC-CD45+. GSEA input was differentially expressed genes ranked by log2 fold change. a) V_VC+CD45+ enriched pathways. b) V_VC-CD45+ enriched pathways (next page). Colour coding for next page is described in the key. FDR <0.25.

b)	VC-CD45+ (Haematopoietic)	
Pathway	NES	FDR
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	-1.76	0.00
HALLMARK_ANGIOGENESIS	-1.58	0.00
HALLMARK_ALLOGRAFT_REJECTION	-1.56	0.00
KEGG_ASTHMA	-1.72	0.00
BIOCARTA_IL12_PATHWAY	-1.70	0.00
HALLMARK_INFLAMMATORY_RESPONSE	-1.49	0.01
KEGG_GRAFT_VERSUS_HOST_DISEASE	-1.66	0.01
REACTOME_PD1_SIGNALING	-1.67	0.01
KEGG_AUTOIMMUNE_THYROID_DISEASE	-1.66	0.01
REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES	-1.67	0.01
REACTOME_COSTIMULATION_BY_THE_CD28_FAMILY	-1.68	0.01
BIOCARTA_TOB1_PATHWAY	-1.66	0.01
KEGG_ALLOGRAFT_REJECTION	-1.65	0.01
KEGG_TYPE_I_DIABETES_MELLITUS	-1.65	0.01
BIOCARTA_NO2IL12_PATHWAY	-1.64	0.01
HALLMARK_KRAS_SIGNALING_UP	-1.45	0.01
REACTOME_GENERATION_OF_SECOND_MESSENGER_MOLECULES	-1.63	0.02
HALLMARK_PANCREAS_BETA_CELLS	-1.42	0.02
HALLMARK_KRAS_SIGNALING_DN	-1.42	0.02
PID_IL12_STAT4_PATHWAY	-1.61	0.02
BIOCARTA_CTLA4_PATHWAY	-1.61	0.02
PID_TOLL_ENDOGENOUS_PATHWAY	-1.61	0.02
HALLMARK_TNFA_SIGNALING_VIA_NFKB	-1.40	0.02
REACTOME_PTM_GAMMA_CARBOXYLATION_HYPUSINE_FORMATION_AND_ARYLSULFATASE_ACTIVATION	-1.60	0.03
HALLMARK_MYOGENESIS	-1.36	0.03
REACTOME_CLASS_A1_RHODOPSIN_LIKE_RECEPTORS	-1.59	0.04
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	-1.58	0.04
HALLMARK_NOTCH_SIGNALING	-1.34	0.04
HALLMARK_COMPLEMENT	-1.34	0.05
REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS	-1.56	0.05
BIOCARTA_COMP_PATHWAY	-1.56	0.05
KEGG_HISTIDINE_METABOLISM	-1.57	0.05
KEGG_HEDGEHOG_SIGNALING_PATHWAY	-1.56	0.05
REACTOME_DEFENSINS	-1.57	0.05
REACTOME_GPCR_LIGAND_BINDING	-1.56	0.05
KEGG_PRIMARY_IMMUNODEFICIENCY	-1.55	0.06
PID_RHODOPSIN_PATHWAY	-1.55	0.06
NABA_SECRETED_FACTORS	-1.54	0.07
BIOCARTA_TOLL_PATHWAY	-1.54	0.07
BIOCARTA_DC_PATHWAY	-1.54	0.07
REACTOME_DOWNSTREAM_TCR_SIGNALING	-1.53	0.07
PID_UPA_UPAR_PATHWAY	-1.53	0.07
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	-1.53	0.07
KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	-1.53	0.07
HALLMARK_INTERFERON_GAMMA_RESPONSE	-1.29	0.08
HALLMARK_COAGULATION	-1.28	0.08
KEGG_PRION_DISEASES	-1.52	0.09
BIOCARTA_NKT_PATHWAY	-1.51	0.09
PID_IL23_PATHWAY	-1.50	0.10
REACTOME_CHONDROITIN_SULFATE_BIOSYNTHESIS	-1.50	0.10
REACTOME_G_ALPHA_I_SIGNALLING_EVENTS	-1.50	0.10
KEGG_HEMATOPOIETIC_CELL_LINEAGE	-1.51	0.10
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	-1.26	0.10
REACTOME_TCR_SIGNALING	-1.50	0.10
KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	-1.50	0.10
ST_G_ALPHA_I_PATHWAY	-1.49	0.11
HALLMARK_XENOBIOTIC_METABOLISM	-1.24	0.12
REACTOME_AMINE_LIGAND_BINDING_RECEPTORS	-1.48	0.12
HALLMARK_IL6_JAK_STAT3_SIGNALING	-1.24	0.12

4.1.5 Renin and Endothelin pathway components are differentially expressed on the haematopoietic and endothelial populations of the dorsal aorta.

The LCM-Seq analysis reported in Chapter 3 revealed that blood pressure regulators renin (*REN)* and endothelin-1 (*EDN1)* are ventrally enriched in the blood vessel wall of the dorsal aorta. As previously mentioned in 4.1.2 ventral enrichment of *EDN1* is also seen in the VC+CD45- endothelial cells. This sorted dorsal aorta RNA-Seq dataset can be used to visualise expression of components of the renin and endothelin pathways in the dorsal aorta and build up the picture of how these cells may interact across sub-populations of the EHT and with the niche. The heatmap in Figure 41 shows normalised expression of some of the main components of these pathways.

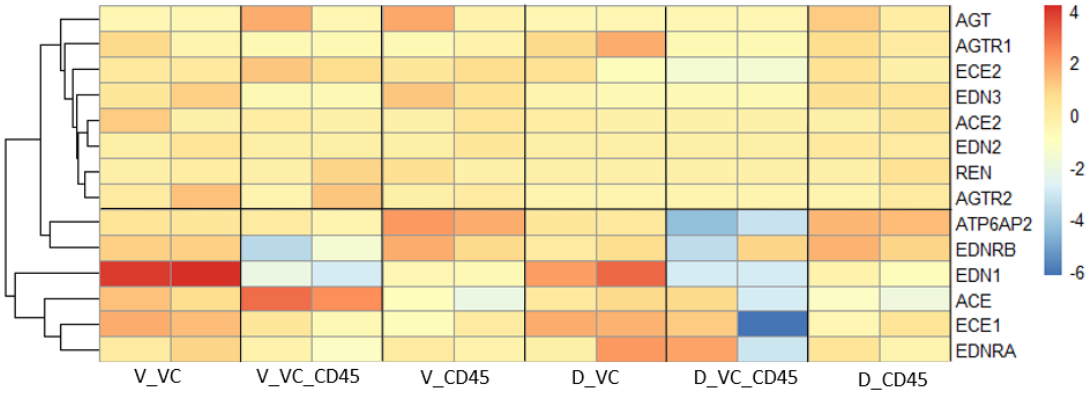


Figure 41. Components of the Renin and Endothelin pathways are differentially expressed on the dorsal-aorta haemato-endothelial populations. Heatmap of normalised expression values for each population. Scale bar shows deviation from the row mean. V = Ventral. D = Dorsal. VC = VC+CD45-, VC_CD45= VC+CD45+. CD45 = VC-CD45+. Some genes are not differentially expressed and are present at low levels across the populations (above horizontal black line). Others show higher levels of expression on specific populations and are not expressed on others (below black line).

REN is not differentially expressed across populations indicating that its strong ventral enrichment is not in the endothelial or haematopoietic populations of the dorsal aorta but may be in the underlying mesenchymal niche cells. *ACE* encoding Angiotensin-converting enzyme, which has been described as an HSC marker in the human AGM (Jokubaitis *et al.*, 2008) is most strongly expressed on the V_VC+CD45+ cells which is the HSC containing population (Ivanovs *et al.*, 2014). *AGT*, which encodes Angiotensinogen the substrate for REN and subsequently ACE is not highly expressed consistently on any population. It was also not expressed in any layer of the LCM-Seq analysis suggesting that, if it is the substrate for REN and ACE in this instance, it derives from another cell source, perhaps through the blood stream. There is no strong expression of either of the angiotensin receptors *AGTR1* or *AGTR2*

on any of the populations. The receptor for prorenin, the inactive precursor to renin, *ATP6AP2*, is expressed most strongly on the VC-CD45+ populations.

EDN1, as described, is expressed across the whole aorta in the VC+CD45- cells but more strongly in the ventral domain. It then has a sharp decrease in expression in the VC+CD45+ population but some lower level expression in seen on the more mature VC-CD45+ population. The gene encoding endothelin-converting enzyme -1, *ECE1*, is also more strongly expressed on the endothelial VC+CD45- cells. Of the endothelin receptors, *EDNRB* rather than *EDNRA* has consistent differential expression between EHT populations with highest expression on the VC-CD45+ populations and lower expression on VC+CD45- cells, with a similar drop in expression to *EDN1* on the VC+CD45+ cells. This suggests a potential autocrine signalling of ET-1 (endothelin-1, protein) via ET-B (endothelin receptor B, protein).

4.1.6 **Modelling interactions of the niche with haematogenic populations of the aorta.**

The LCM-Seq spatial transcriptome and the sorted dorsal aorta populations analyses, can be brought together to model how niche signals may interact with the endothelial and HSPC cells to promote HSC emergence and support maturation. To this end, LCM-Seq ventrally enriched secreted factors and receptors from the VC+CD45- vs. VC+CD45+ populations were analysed for known interactions using the STRING database (Fig.42a (V_VC+CD45+), b (V_VC+CD45-)). The number of connections corresponds to the size of the node on the interaction networks potentially indicating top of the hierarchy signalling factors. Secreted factors are in red, receptors are in blue. An interaction could be one of the following; activation, inhibition, binding, catalysis, phenotype, posttranslational modification, reaction, transcriptional regulation (see Materials and Methods). *SPP1* which encodes osteopontin, is a highly significant ventralised secreted factor in the Mid domain which directly binds the top 2 most significantly differentially expressed receptors on the V_VC+CD45+ HSPC population compared to the V_VC+CD45- endothelial population, CD44 and ITGA4 (Fig.42a) highlighting a potential niche-HSPC signalling interaction. However, *SPP1* is also expressed on the HSPC population (but not VC+CD45- endothelial) itself so signalling with CD44 may be autocrine.

The secreted factor with the highest amount of connections to HSPC receptors is *PLAU*, encoding the serine protease urokinase that converts plasminogen to plasmin (O'Leary *et al.*, 2016) and interacts with numerous diverse receptors types such as the integrin receptors ITGAM and ITGB2, the TNFα receptor TNFRSF1B, the G protein-coupled receptor GPR84 and the immunoglobulin receptors LILRB2, FCER1G, CD300A and LAIR1. PLAU modulates cell adhesion through the integrins facilitating cell mobility during migration raising the idea that this niche-HSPC interaction aids the movement of HSPCs as they round up and bud into the lumen, detaching from the endothelial and ECM structures of the vessel wall. Furthermore, *PLAU*, along with *EDN1* and the smooth muscle constrictor *NMB* all interact with *PTAFR* (Platelet activating receptor) which is also a potent smooth muscle constrictor indicating that these niche and HSPC signals may act in concert to regulate vascular tone and constriction.

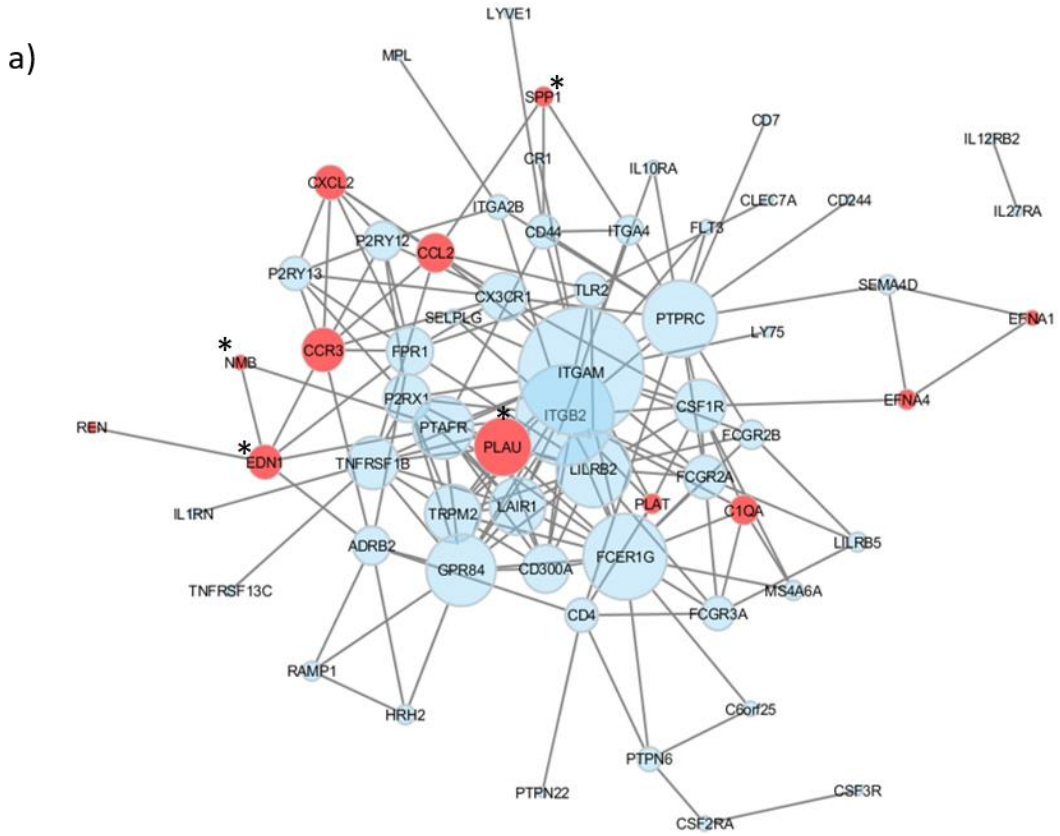
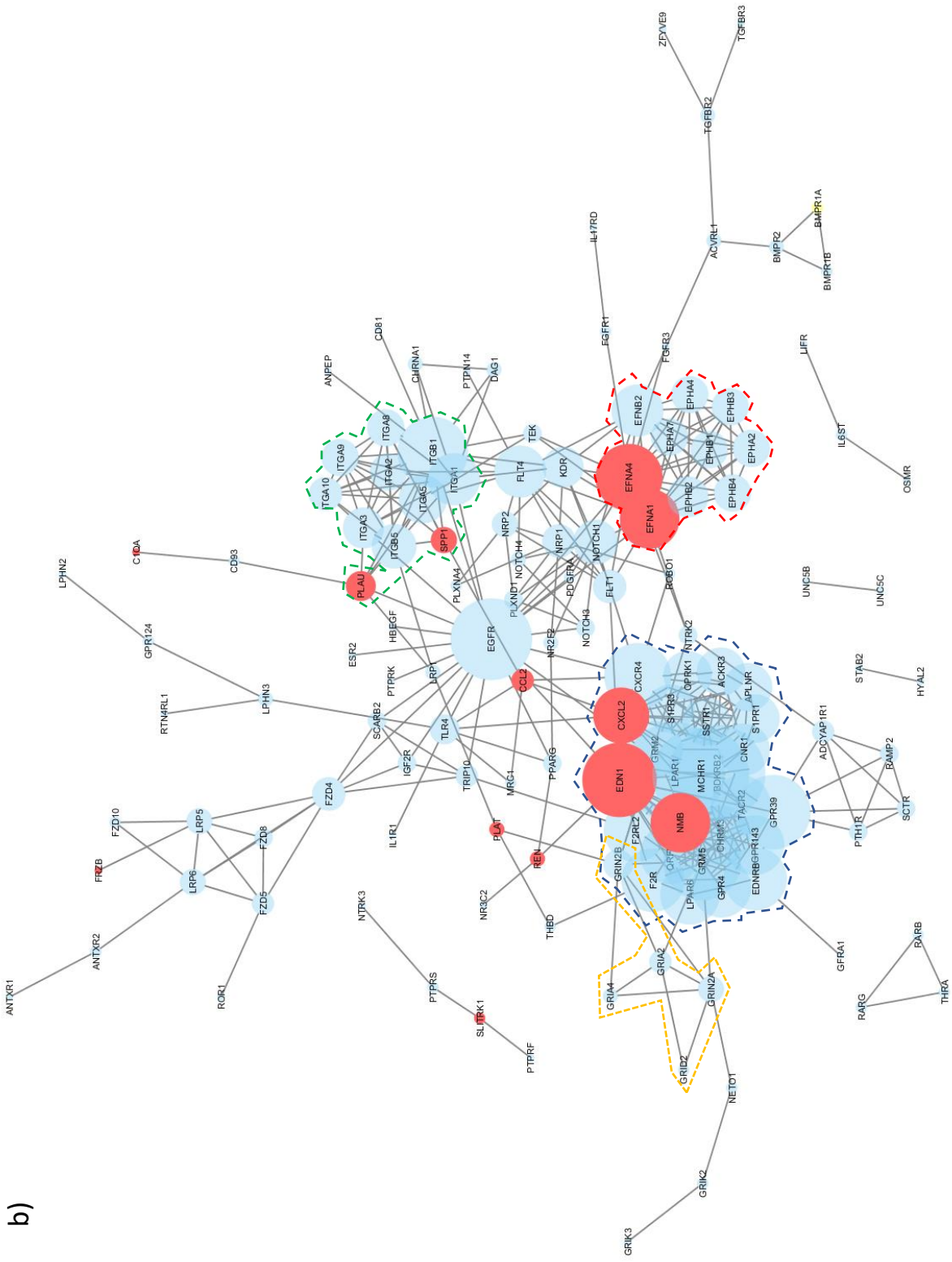


Figure 42 . Modelled interactions of the niche with the ventral endothelial and HSPC population. StringDB interactome of ventrally enriched secreted factors (Inner and Mid from LCM-Seq) in red with a) V_VC+CD45+ or b) V_VC+CD45- (next page) enriched receptors in blue. High confidence interactions only (>0.7/1). Asterisk highlights secreted factors discussed in results. Size of the node reflects number of interactions. b) Clusters are grouped together by coloured dotted lines; green = integrin receptors, red = ephrins and ephrin receptors, blue = G-protein coupled receptors, yellow = neuroactive ligand receptors.


b)

In the network of niche secreted proteins and VC+CD45- endothelial receptors (Fig.42b) *EDN1* has the highest number of connections indicating a central signalling role to the endothelial cells, although as the VC+CD45- express *EDN1* highly themselves (Fig.36), this is likely to be in part, if not fully, autocrine signalling. A direct action of *EDN1* on this population is further evidenced by expression of endothelin receptor *EDNRB* as previously mentioned (Fig.41). In contrast, *NMB* which also has a high number of overlapping connections with *EDN1*, is highly expressed in the ventral sub-aortic mesenchyme (Mid) indicating a paracrine mode of action. The shared connections cluster into a group of G-protein coupled receptor (GPCR) signalling related genes. This includes further blood pressure regulation receptors *BDKRB2* (B2 bradykinin receptor) and *APLNR* (Apelin receptor). Taken together, components of systems involved in regulating vascular constriction and dilation are differentially expressed in the endothelial, HSPC and mesenchymal cells indicating a complex, compartmentalised regulation of vascular tone.

PLAU and *SPP1* cluster on the VC+CD45- network with integrin receptors as with VC+CD45+ indicating that they may influence focal adhesion and cell migration at both the VC+CD45- and VC+CD45+ level on different receptors. One striking difference between VC+CD45- and VC+CD45+ types of receptors are the respective presence and absence of ephrin receptors highlighting a downregulation of these receptors during HSPC generation. The ventral sub-aortic mesenchyme secreted ephrins *EFNA1* and *EFNA4* have many interactions with this group of receptors as expected. Several neuroactive ligand receptors are also upregulated in the VC+CD45- vs VC+CD45+ cell group, although they have minimal connections to the niche secretome captured with LCM-Seq.

TGFβ signalling is a ventrally enriched pathway in both the endothelial, perivascular and sub-aortic mesenchyme compartments (VC+CD45-, Inner, Mid resepectively). Notably the VC+CD45- endothelial cells have several receptors to the TGFβ ligand superfamily including BMP receptors *BMPR2, BMPR1A, BMPR1B and ACVRL1 and* TGFβ receptors *TGFBR2, TGFBR3 and ZFYVE9* (Fig.42b) demonstrating that the VC+CD45- are effector cells for this signalling pathway. TNFα signalling was also ventrally enriched in the perivascular and sub-aortic mesenchyme but enriched on the HSPC over the endothelial compartment. However, there are differing TNFα signalling receptors enriched on both the HSPC cells (TNFRSF13C, TNFRSF1B) (Fig.42a) and endothelial cells (TNFRSF19, TNFRSF10D, TNFRSF12) (Fig.42b) demonstrating that the effects of this signalling pathway on the different cell types are modulated through different receptors.

Although secreted factors in the niche are just one way in which it may influence the cells of the dorsal aorta, this analysis highlights potential interactions of the niche with the endothelium or HSPC populations for further investigation.

4.2 Single cell RNA-Seq analysis of the ventral dorsal aortic cell populations

RNA-Seq of bulk sorted populations is useful for modelling interactions of the niche with the populations along the EHT axis as well as the dynamics in gene expression during this process. However, the sorting based on VC and CD45 expression does not allow for the exploration of transcriptional differences in sub-populations. For example, the VC+CD45- endothelium may contain both non-haematogenic and haematogenic endothelium or pre-HSCs that do not yet express CD45+. Although the V_VC+CD45- vs D_VC+CD45- analysis provides evidence of potential genes that may confer haematogenic vs non-haematogenic identity, the signals that mark the initiation of HSC emergence from the endothelium may derive from relatively small sub-populations and therefore get lost in the bulk population transcriptome. To address this, 10X single cell sequencing of the ventral domain of the dorsal aorta was carried out. The dorsal aorta of a CS16 specimen was sub-dissected into ventral (AoV) and dorsal (AoD) domains, dissociated into single cells and CD34+ cells were purified using MACS column separation (Fig.43). CD34+ cells mark both the endothelial cells and the IAHCs. Clustering of these cells into sub-populations would thus allow resolution of the heterogeneity of the ventral endothelium and haematopoietic progenitor cells. Furthermore, lineage inference tools would facilitate identification of the direct endothelial ancestors of HSPC cells.

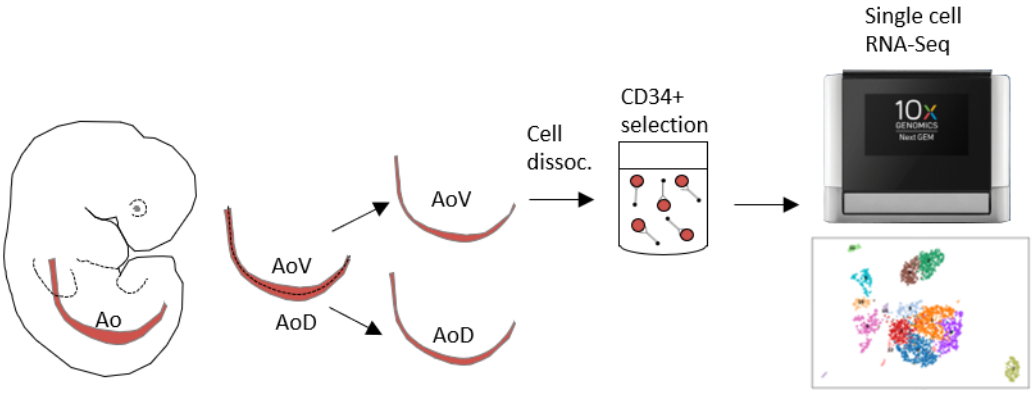


Figure 43 Experiment plan for single cell analysis of CD34+ cells in the ventral dorsal aorta. The dorsal aorta (Ao) was sub-dissected by hand into ventral (AoV) and dorsal (AoD) regions. The AoV was dissociated and CD34+ cells were selected using MACS bead column separation. 10X genomics was used to prepare the single cell RNA-Seq libraries for sequencing and subsequent analysis.

4.2.1 Quality control testing of the dataset and removal of confounding effects

The AoV CD34+ cells were run on two lanes of the 10X chip resulting in a combined total of 2,421 cells with 71,270 mean reads per cell. The data were pre-processed based on widely accepted quality control metrics (Ilicic *et al.*, 2016).The number of genes per cell, the number of unique feature counts per cell (n_counts) and the percentage of mitochondrial genes were visualised using violin plots

(Fig.44a). Cells were removed that had above a threshold of 3500 unique genes detected, determined by the graph

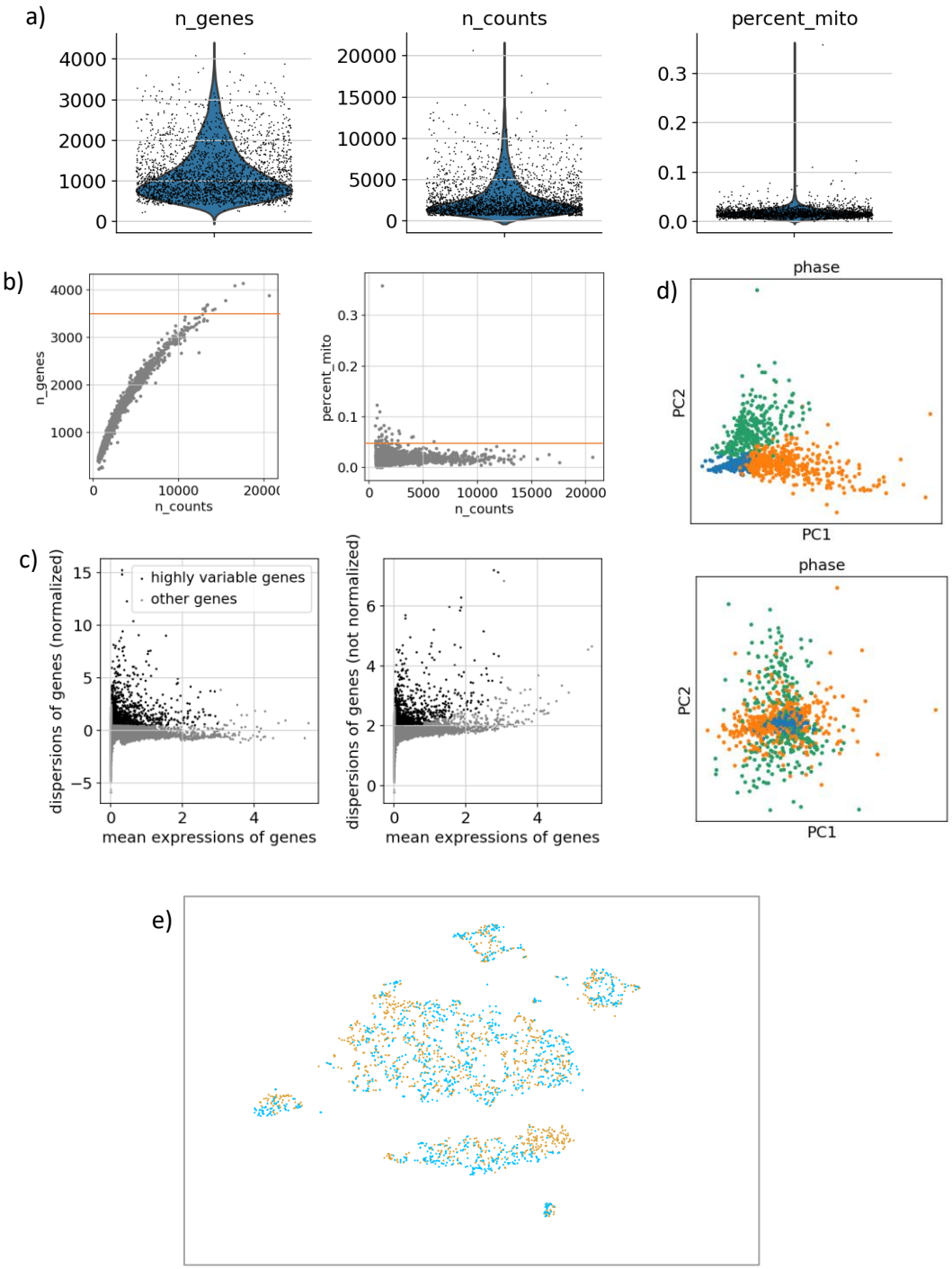


Figure 44. Quality control testing of the CS16 AoV CD34+ cell 10X single cell dataset. a) Violin plots showing the number of genes (n_genes), number of UMI counts (n_counts) and percentage of mitochrondrial genes in total (percent_mito) per cell. b) Dot plots showing cut-off thresholds (red line) for n_genes and percent_mito. Cells above threshold were discarded from the analysis. c) Normalised and non-normalised dispersions of genes, indicating in black the highly variable genes used in the differential expression testing. d) Principal component plots for cells based on cell cycle gene expression before (upper) and after (lower) regression of cell cycle effects. f) Overlay of tSNE clustering from two technical replicates shown in different colours.

of n_counts vs. n_genes, as high numbers indicate cell doublets or multiplets (Fig.44b). As high mitochondrial reads are evidence of low-quality cells, cells with more than 0.05% mitochondrial genes were also removed. The numbers of UMI counts per cell had a large range of ~2,000-10,000 counts with a median of 2,768. The reads per cell were then normalised so that counts become comparable among cells.

Highly variable genes were identified (Fig.44c) and the data were subsetted to include only these as this enhances the discovery of biological variation between cells in downstream analyses (Brennecke *et al.*, 2013). Furthermore, PCA analysis of cell cycle genes in the dataset highlighted that some of the variance in the dataset is explained by cell cycle so the effect of these genes was regressed out (Fig.44d). As the cells were run on two lanes of the 10X chip, they were checked for batch effects. Basic t-SNE clustering was done on the Loupe Cell Browser software and the cells were colour coded for each lane (Fig.44e). There was an even distribution of cells from both lanes in each cluster indicating that there was no large batch effect.

4.2.2 Dimension reduction of the single cell dataset reveals haematopoietic, endothelial and niche sub-populations.

Clustering the cells was done using a nonlinear dimensionality-reduction technique, uniform manifold approximation and projection (UMAP) using the louvain algorithm which generated 10 clusters (Fig.45a). By visualising patterns of expression of gene markers some basic cell-type identities could be conferred to the clusters. *CD34* and *CDH5* (VC) expression indicates that the endothelial cells and their immediate derivatives expressing these markers include clusters (Cl.) 0, 1, 2, 5, 7, 8, 10 indicating substantial heterogeneity across these populations in the dorsal aorta. The additional clusters most likely derive from contaminating CD34- cells from the niche. Cl.3 expresses *PDGFRA* and *ACTA2* identifying it as containing mesenchymal cells whereas clusters 4 and 9 express *EPCAM* and therefore represent an epithelial fraction. *PTPRC* (CD45) expression is highest on cl.6. There are only a few scattered *CD34* and *CDH5* expressing cells in this cluster indicating that this contains mainly more mature haematopoietic cells – however, the possibility that *CD34* or *CDH5* expression levels were just below the detection threshold cannot be ruled out. *PTPRC* is also expressed on a fraction of cl.5 indicating that this population contains the double positive VC+CD45+ HSPC population and puts cl.5 as the most likely candidate for endothelial cells committed to a haematopoietic fate. Cl.10 expresses the erythroid marker *GYPA* encoding Glycophorin-A indicating that these VC+GYPA+ cells are erythroid progenitors.

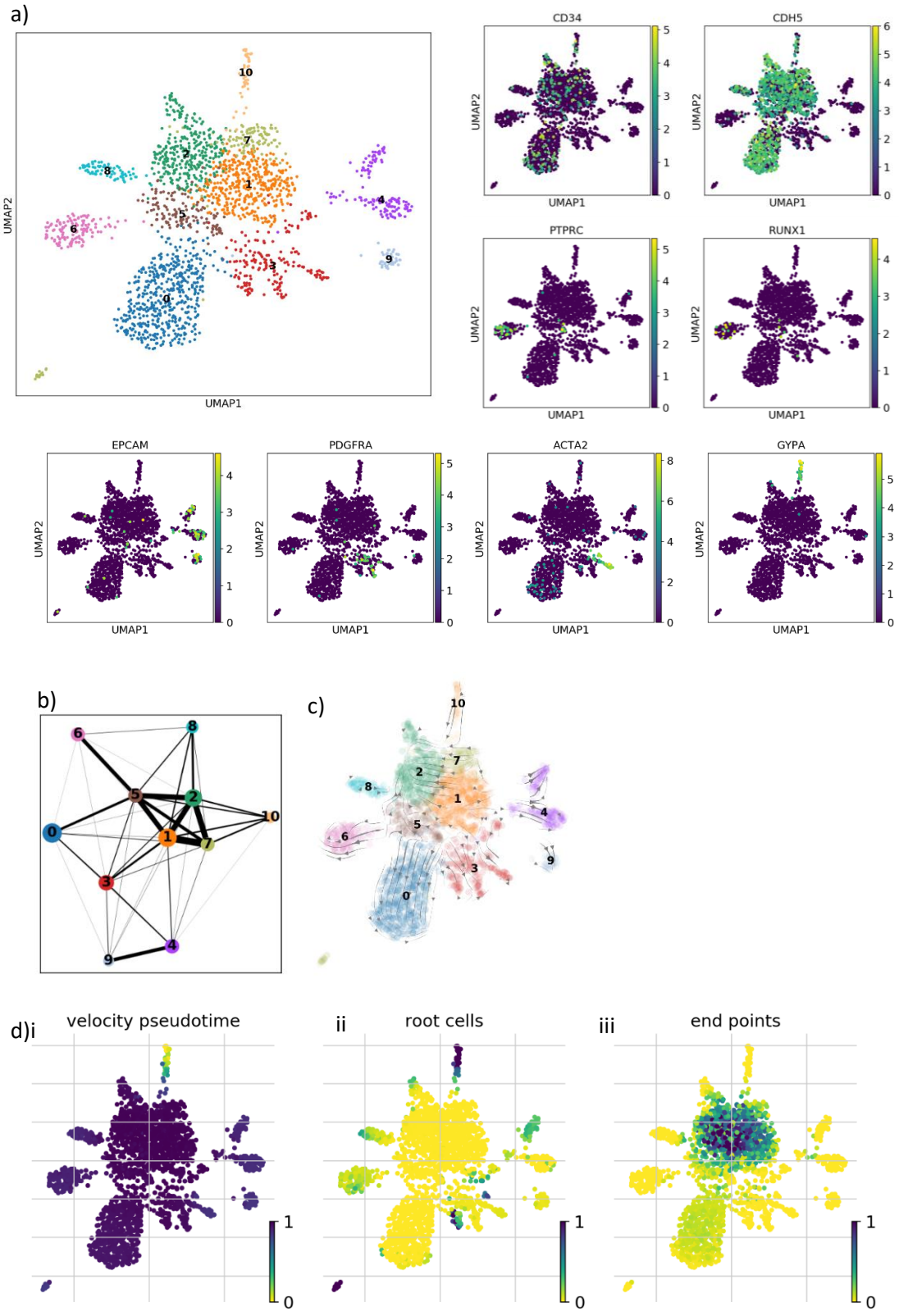


Figure 45. UMAP clustering reveals sub-populations of the dorsal aorta. a) UMAP partitioning and Louvain cluster identification reveals 10 clusters. Expression of known markers of endothelial, haematopoietic and mesenchymal populations are shown in smaller plots. b) PAGA indicates relatedness of clusters. c) RNA Velocity mapping of differentiation trajectories. d)i. RNA Velocity pseudotime trajectory plot with separate plots indicating ii. root cells and iii. end points.

4.2.3 Cluster inter-connectivity and RNA Velocity mapping reveals potential lineage trajectories.

The first step towards lineage trajectory inference is to define connectivity between the clusters. For this, Partition-based graph abstraction (PAGA) was used (Wolf *et al.*, 2019). This first uses PCA dimensional reduction of the data to determine neighbourhood distances between cells. The data are then partitioned using the Louvain algorithm as used before for generating the UMAP. Nearest neighbour connections between clusters are weighted statistically based on the number of connections as a fraction of the connections expected if assigned randomly.

The resulting PAGA graph shows high inter-connectivity between endothelial clusters 5,1,2 and 7 with several clusters branching off from them (Fig.45b). Cl.6 has a weighted connection with cl.5 lending further evidence that these haematopoietc cells derive from the cl.5 endothelial cells. Intriguingly, the large endothelial cl.0 is apart from the central network of endothelial cells connecting most strongly with cl.5. Cl.10 and cl.8 *CDH5* expressing cells also branch off from the central network. The mesenchymal cl.3 connects with endothelial clusters 1 and 0, albeit with lower statistical weighting, which raises the question of whether there is a shared ancestor between the two cell types. The EPCAM expressing clusters have a weighted connection suggesting that they are related cell types.

RNA velocity, which uses the relative proportions of spliced and unspliced transcripts to predict the direction of cellular state progression (Manno *et al.*, 2018), was used to infer a lineage trajectory from the data (Fig.45c) as well as predict cells at the root of the trajectory and at the end (Fig.45di-iii). Clusters 2 and 1 of the endothelial cells were determined as 'end points' thus indicating a stable population of non-differentiating cells. The velocity graphs indicate a movement from cl.5 towards the haematopoietic population (cl.6) as well as towards cl.0 suggesting a bifurcation in cell fate potential of cl.5 to transition to haematopoiesis or to an alternative endothelial cell type. Whereas the velocity of most clusters is moving away from the central endothelial clusters, cl.10 is determined as containing root cells with velocity towards the endothelial cells, which unusually suggests that they are differentiating towards an endothelial fate and may represent a population of endothelial progenitors. Indeed, this was the only clear pseudotime trajectory that the RNA Velocity tool could define (Fig.45d). Cl.3, has 2 groups of cells highlighted as being root cells indicating that there are alternately differentiating sub-groups within this cluster. The mesenchymal markers *PDGFRA* and *ACTA2* mark different cells within that cluster suggesting it may be made up of multiple mesenchymal cell types.

4.2.4 A more in-depth analysis of cluster-specific cell identities

Heterogeneity within the endothelial populations

In order to better define the cell-types in each cluster a Wilcoxon rank-sum test was used to find differentially expressed genes in each cluster compared to the rest. A dot plot identifying the top 5

markers for each cluster are shown in (Fig.46a) with a separate heatmap of the top 50 genes (Fig.46b) for each cluster. Furthermore, the top 100 genes per cluster were computed for overlaps with Hallmark pathways (Fig.47). Clusters 1 and 2 make up the bulk of the central CD34+ endothelial group of clusters. The heatmap shows that these clusters are highly similar to each other as well as the other endothelial clusters, 5, 7, and 8 and with even some of the top markers for each cluster shared among them (Fig.46).

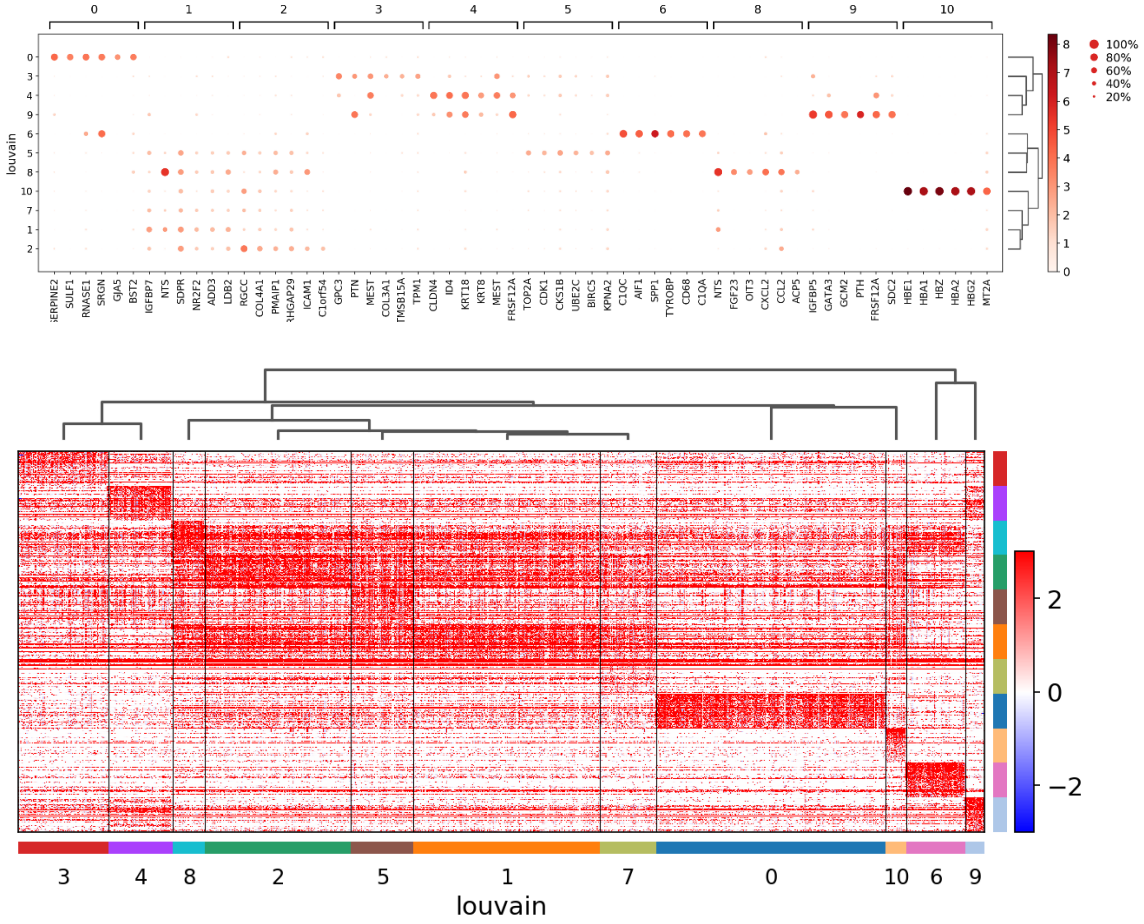
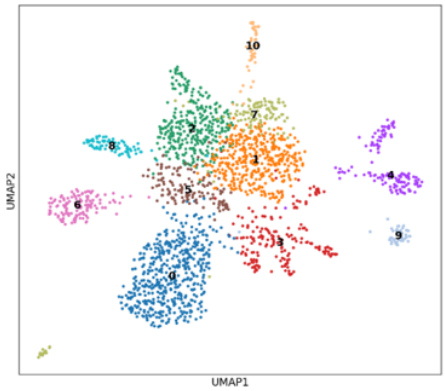


Figure 46. Heatmap visualisation of AoV single cell dataset Louvain cluster relatedness. a) Dot plot of top 5 markers for each cluster. b) Heatmap of expression for top 50 genes per cluster with hierarchical clustering. The close relatedness of endothelial clusters 8,2,5,1 and 7 can be visualised as well as the distinct signature of endothelial cluster 0.



CDH5+ Endothelial	0	HALLMARK_TNFA_SIGNALING_VIA_NFKB	3.7 e^{-19}
		HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	3.79 e^{-16}
		HALLMARK_KRAS_SIGNALING_UP	1.01 e^{-14}
	1	HALLMARK_MYC_TARGETS_V1	3.63 e^{-3}
		HALLMARK_ESTROGEN_RESPONSE_LATE	1.52 e^{-2}
		HALLMARK_MYOGENESIS [200]	1.52 e^{-2}
	2	HALLMARK_TNFA_SIGNALING_VIA_NFKB	1.49 e^{-17}
		HALLMARK_APOPTOSIS	5.73 e-8
		HALLMARK_APICAL_JUNCTION	2.1 e-7
	5	HALLMARK_E2F_TARGETS	1.06 e^{-41}
		HALLMARK_G2M_CHECKPOINT	5.45 e^{-38}
		HALLMARK_MYC_TARGETS_V1	3.63 e^{-23}
	7	HALLMARK_MYC_TARGETS_V1	1.22 e^{-8}
		HALLMARK_MITOTIC_SPINDLE	1.66 e^{-2}
	8	HALLMARK_MTORC1_SIGNALING	3.7 e^{-19}
		HALLMARK_HYPOXIA	1.51 e^{-14}
		HALLMARK_UNFOLDED_PROTEIN_RESPONSE	1.16 e^{-12}
	10	HALLMARK_HEME_METABOLISM	1.12 e^{-44}
		HALLMARK_KRAS_SIGNALING_UP	8.95 e^{-4}
		HALLMARK_MYC_TARGETS_V1	8.95 e^{-4}
CD34+ ACTA2+	3	HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	1.82 e^{-26}
		HALLMARK_UV_RESPONSE_DN	2.01 e^{-8}
		HALLMARK_E2F_TARGETS	2.86 e^{-6}
CD34- CD45+	6	HALLMARK_TNFA_SIGNALING_VIA_NFKB	6.53 e^{-16}
		HALLMARK_COMPLEMENT	1.62 e^{-11}
		HALLMARK_IL6_JAK_STAT3_SIGNALING	1.7 e^{-5}
CD34- EPCAM+	4	HALLMARK_GLYCOLYSIS	4.52 e^{-7}
		HALLMARK_HYPOXIA	2.47 e^{-6}
		HALLMARK_MTORC1_SIGNALING	2.47 e^{-6}
	9	HALLMARK_OXIDATIVE_PHOSPHORYLATION	5.65 e^{-32}
		HALLMARK_CHOLESTEROL_HOMEOSTASIS	1.35 e^{-5}
		HALLMARK_ESTROGEN_RESPONSE_EARLY	5.76 e^{-4}

Figure 47. Computed overlaps of cluster marker genes with Hallmark gene sets resolves niche signalling to distinct cell populations. Table of top 3 HALLMARK pathways for each cluster.

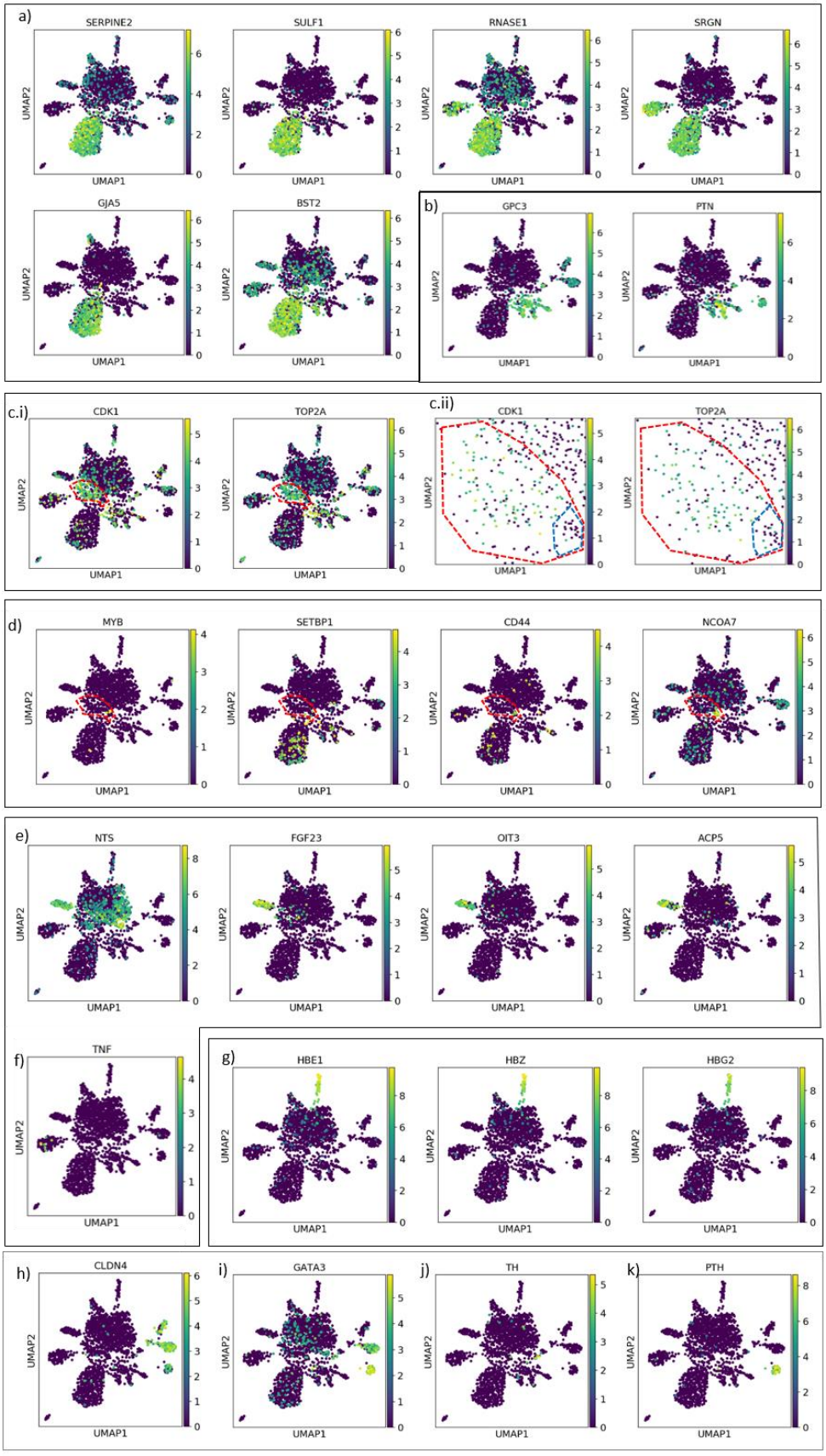

a)
SERPINE2
SULF1
RNASE1
SRGN
GJA5
BST2
b)
GPC3
PTN
c.i)
CDK1
TOP2A
c.ii)
CDK1
TOP2A
d)
MYB
SETBP1
CD44
NCOA7
e)
NTS
FGF23
OIT3
ACP5
f)
TNF
g)
HBE1
HBZ
HBG2
h)
CLDN4
i)
GATA3
j)
TH
k)
PTH
UMAP1
UMAP2

Figure 48 Expression maps of selected cluster marker genes onto the AoV single cell dataset for a) cluster 0 b) cluster 3, c)i 'haematogenic' cluster 5 and ii. close view of area enclosed by dotted red line in i. d) Expression of VC+CD45+ marker genes from bulk sorted EHT population dataset. e) marker genes for cluster 8, g) cluster 10 h) cluster 4, i) and k) cluster 9. f) Expression of *TNF* localised to cluster 6A m. j) Expression of *TH* localised to a region of Cluster 3. Cluster boundaries are defined in Fig.44a.

Cluster 0: A distinct endothelial sub-type

By contrast, the endothelial cl.0 has a very different signature to the rest of the endothelial clusters indicating that it represents a distinct sub-type of endothelial cells. Top marker genes for this are the Heparan Sulfate Proteoglycan *HSPG2*, the heparin sulfate endosulfatase *SULF1* and Gap Junctiom Protein Alpha 5 (*GJA5*) (Fig.46a,48). It also has an enrichment for 'TNFα signalling via NFκB' (Fig.47) and expresses several pathway components such as *EDN1, PTGS2* and the transcription factors *JUNB* and *FOS*. This pathway was ventrally enriched in the Inner and Mid layers of the LCM-Seq analysis and in the bulk-population analysis, was shown to increase along the EHT axis. TNF itself is expressed at highest levels on the VC-CD45+ bulk-seq population and the *PTPRC* (CD45) expressing population (cl.6) in the single cell dataset (Fig.48f). It is tempting to postulate that the TNFα signalling via NFκB pathway could be originating from the haematopoietic cells and signalling back towards the niche (see discussion). Cl.0 also has enrichment of Notch pathway components including the ligands *DLL4, JAG1* and the effector TF *HEY2.*

Cluster 5: Proliferating endothelial cells

Cl.5 which contains the HSPC population and is lineage linked to cl.6 haematopoietc population shares expression of many genes with the other endothelial clusters. However, cl.5 is additionally enriched for genes involved in mitosis such as the mitosis cyclin *CDK1*, and *TOP2A*, a DNA topoisomerase involved in chromatid separation (Fig.46a,48c). Furthermore, highly enriched gene sets include proliferation related pathways such as 'E2F Targets', 'G2M Checkpoint', and 'MYC Targets V1' (Fig.47). This suggests that the cl.5 endothelial cells are related to the bulk of the endothelial cells but with a higher proliferative status. The marker genes however are noticeably down regulated in the cl.5 subpopulation of the cluster expressing *PTPRC*. Markers from the bulk-seq VC+CD45+ population were mapped on to the single cell dataset and although their expression was not restricted to cl.5, their expression within cl.5 was restricted to the PTPRC+ portion (Fig.48d) This indicates that the majority of cells within cl.5 are not HSPCs but are mostly proliferating endothelial cells. Cl.5 lacks any discerning markers or transcription factors that distinguish it from the other central endothelial clusters (cl.1 and cl.2) indicating that they are a sub-population of the endothelial cells with a higher proliferative status. This could perhaps indicate that they are endothelial progenitors. This proliferation signature is then absent in the cells which express *PTPRC* and other HSPC signature genes indicating that proliferation is slowed as a haematopoietic identity is acquired.

Cl.8: Hypoxia sensitive endothelial cells

Cl.8 is another endothelial population branching away from the central endothelial group of clusters. Although, it has a more similar transcriptional signature to the central group than cl.0 it does express additional markers including neurotensin, *NTS*, fibroblast growth factor 23, *FGF23*, Oncoprotein Induced Transcript 3, *OIT3* and the osteopontin regulator, acid phosphatase 5, *ACP5* (Fig.46a, 48e). Cl.8 has enrichment for 'MTORC1 Signalling' and 'Hypoxia Signalling' which are both pathways found in the ventral niche using LCM (Fig.47).

Cl.10: ***Embryonic/foetal erythromyeloid progenitors***

The final cluster branching off from the central endothelial group of clusters is cl.10 which expresses haemoglobin genes (Fig.46a,48g). Intriguingly, these include Epsilon, Zeta and Gamma haemoglobin subunits *HBE1*, *HBZ* and *HBG2* which are embryonic and foetal haemoglobins respectively. As these earlier erythroid progenitors are generally considered to emerge in the yolk sac it is surprising to find them seemingly related to the bulk of the endothelial cells in the AoV. It is also unusual that their velocity is moving towards the main bulk of the endothelial cells as if they are differentiating from an erythroid towards an endothelial lineage rather than vice versa.

Marker genes for cl.7 are mitochondrial genes indicating that these are low quality cells that were below the percentage mitochondrial genes per cell threshold and so evaded being filtered out.

Cl.3: mesenchymal/perivascular cells

A top marker for cl.3, which was identified as a mesenchymal/perivascular cluster based on *ACTA2*, *PDGFRA* and *PDGFRB* expression, is *GPC3* which encodes glypican 3, a cell surface proteoglycan to which heparan sulfate chains attach (Fig.46a, 48b). It has a described role in inhibition of Hedgehog signalling and activation of Wnt signalling, both of which are pathways found in the niche (Hedgehog on ventral VC+CD45- and VC-CD45+ cells, Wnt signalling in the Inner and Mid layers vs Outer)(Capurro *et al.*, 2008; Capurro *et al.*, 2014). *PTN*, Pleiotrophin, is another top marker which is a secreted factor that binds heparin to regulate many processes including cell migration, growth, differentiation, proliferation and survival and, moreover, is important for haematopoietic regeneration in the bone marrow(Himburg *et al.*, 2018; Deuel *et al.*, 2002). Taken together, this indicates that this cluster contains important cells that regulate various biological processes in the AGM niche. The overlap in heparan proteoglycan related genes with cl.0 may also indicate some relatedness between the clusters or shared signalling and functionality. Furthermore, both clusters 3 and 0 have a strong enrichment for the 'Epithelial Mesenchymal Transition' gene set suggesting they influence cell mobility and migration (Fig. 47).

Cl.4 and 9: Epithelial cells

Clusters 4 and 9 express *EPCAM* which indicate that they are neither endothelial nor perivascular (Fig.45a). The top marker for cl.4 is *CLDN4*, a claudin that is part of epithelial cell tight junctions indicating that these cells derive from an epithelial structure (Fig.46a, 48h). Cl.9, has high expression of *GATA3*, which suggested that they might be sympathoadrenal cells (Fitch *et al.*, 2012)(Fig.48i). However, the sympathoadrenal marker TH was not co-expressed on this cluster but on a sub-population of cl.3 (Fig.48j). There is some overlap in gene expression between clusters 4 and 9. *GATA3* is also expressed on one branch of cl.4 and cl.9 expressed *CLDN4* at a lower level (Fig.48h, i). Cl.9 does however exclusively express parathyroid hormone (*PTH*) indicating that they are endocrine cells (Fig. 48k). Intriguingly, cl.4 enriched HALLMARK pathway are similar to cl.8, including mTORC1 signalling and hypoxia (Fig.47), potentially indicating that these clusters may co-operate. Glycolysis pathway and hypoxia in cl.4 and oxidative phosphorylation pathway enrichment in cl.9 indicates both anaerobic and aerobic respiration in the niche.

Taken together this dataset demonstrates how the AoV endothelial cells are heterogenous with several resolved cell types which may have specialised roles. It also suggests that endothelial cells destined to become haematopoietic may increase their proliferative status before subsequent acquisition of haematopoietic identity. Finally, it also provides additional evidence of overlap in signalling pathways between the endothelial cells and other niche cells reinforcing the idea of signalling interplay between the niche and aorta.

4.2.5 Resolution of niche secreted factors to cell sub-populations

As the single cell dataset includes both endothelial and non-endothelial cells from the AoV, expression of ventrally enriched secreted factors from the LCM-Seq Inner and Mid domains can be resolved to clusters and their cell origin revealed (Fig.49). Several of these are expressed on the endothelial clusters themselves, such as *CCL2, EDN1, EFNA1* and *CXCL2* although expression levels differ. *EDN1* is most highly expressed on cl.0 and *CCL2* and *CXCL2* have highest expression on cl.8. *PLAT* has notably higher expression on the EPCAM+GATA3+PTH+ cl.9. The Complement C1qA Chain gene *C1QA* unexpectedly is most highly expressed on the haematopoietic cl.6. However, cl.6 also has the highest expression of osteopontin, *SPP1* suggesting that expression of *SPP1* in the ventral mesenchyme may derive from macrophages in the sub-aortic mesenchyme. The Wnt signalling modulator *FRZB* is most highly expressed in the cl.3 mesenchymal cells. This corresponds with the Wnt signalling activator *GPC3* being the top marker for this cluster.

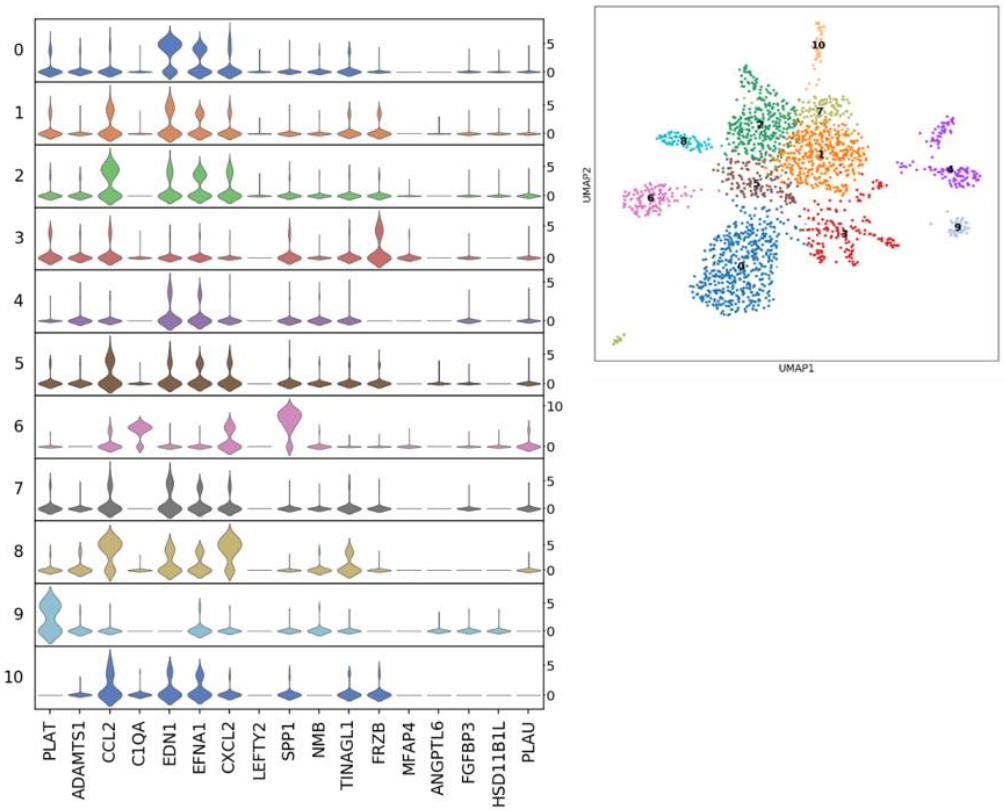


Figure 49. Resolution of niche secreted factors to cell sub-populations. Violin plots showing expression of secreted factors in the Inner and Mid ventral domains from the LCM-Seq analysis for each cluster (1-10). Violin plot colours correspond to Louvain cluster colours.

4.2.6 Resolution of Renin and Endothelin pathway components to sub-populations

EDN1 and *REN* were highly ventrally enriched in the sub-aortic mesenchymal cells from the LCM-Seq analysis. The single cell analysis was used to map the main components of their signalling pathways to resolve their expression to cell populations (Fig.50). In the bulk analysis, *EDN1* was found to be most highly expressed in the ventral VC+CD45- endothelial population. The single cell analysis enabled further resolution of *EDN1* expression; it is found expressed across all the endothelial cells but at a higher and more consistent level in cl.0 (Fig.50a). Resolution of *ECE1* to cl.0 was even more pronounced providing further evidence that this cluster has the most active endothelin signal. Cl.0 also has enrichment of 'TNFα signalling via NFκB' pathway, to which *EDN1* contributes (Fig.47). *EDN1* was also enriched in the LCM-Seq V_Mid (vs. D_Mid) region which may derive from additional expression in the *EPCAM*+ cl.4 and low expression on mesenchymal cl.3 (Fig.50a) As in the bulk-seq analysis, *EDNRB* is expressed on the CD45+ haematopoietic cells in cl.6 but also on the endothelial cells coincident with hotspots of *EDN1* expression. *EDNRA* was however, expressed exclusively on the

perivascular/mesenchymal cl.3. This suggests a two-direction mode of action of *EDN1* from the endothelium signalling towards the haematopoietic cells in the lumen via *EDNRB* and towards the perivascular cells via *EDNRA*.

REN and *ACE* themselves were not detected in this dataset (see discussion) and *AGT* and its receptor *AGTR1* were very lowly expressed in just a few cells (Fig.50b). *AGTR2* expression was confined to cl.0 suggesting that both renin and endothelin pathways act primarily through the same subset of endothelial cells. The renin receptor *ATP6AP2* was expressed broadly across the whole dataset but strongly upregulated on cl.9.

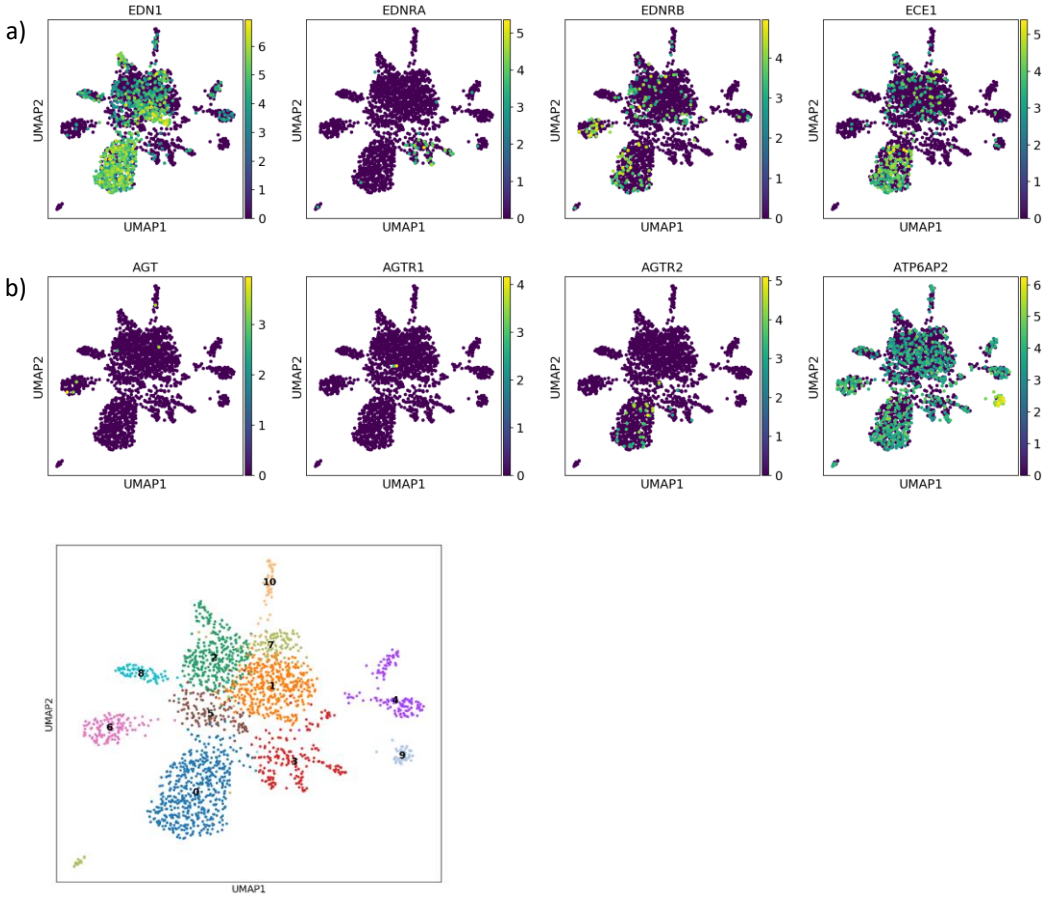


Figure 50. Resolution of Renin and Endothelin pathway components to sub-populations. Gene expression was mapped for components of the Endothelin (a) and Renin (b) pathways. *REN* and *ACE* were not expressed in this dataset.

4.3 Immunohistochemical and RNA Scope visualisation of Renin and Endothelin gene and protein expression in the CS16 AGM

4.3.1 EDN1 expression

The multiplex fluoresecent in situ hybridisation technique RNAScope was used to further spatially resolve expression of *REN* and *EDN1* in the AGM. *EDN1* was confirmed as being expressed in the aortic endothelium across the whole D-V axis (Fig.51a-c). However, there are endothelial cells and groups of cells which had a higher density of *EDN1* probe signal to make a signal cluster, which will be referred to as *EDN1* hotspots. ImageJ was used to define these hotspots as being larger than 300 pixels and the numbers of hotspots were counted in the ventral and dorsal domains of aortas from 14 tissue sections spanning the embryo from the hind gut to the lungs (Fig.51d). Significantly more hotspots were found in the ventral domain compared to the dorsal (paired t-test, $p<0.01$) which confirms the ventral enrichment of endothelial *EDN1* found in the bulk-seq analysis (Fig.51dii). Furthermore, *EDN1* hotspots were found underlying IAHCs (Fig.51b, c). To formally quantify a correlation between position of *EDN1* hotspots and IAHCs sister sections were stained with VC+, CD45+ and Runx1+ to mark the position of IAHCs and Runx1+ endothelial cells (e.g. Fig.51a, b). Their positions were then correlated with the positions and frequencies of *EDN1* hotspots on the sister sections. There was a strong positive correlation between the number of Runx1+ endothelial cells/IAHCs in a dorsal aorta section and the presence of an *EDN1* hotspot in a proximal position on an adjacent section (Fig.52a, $R = 0.9$, p=1.4e-05) indicating that most Runx1+ IAHCs are found near higher local *EDN1* concentrations. Vice versa, there was a positive correlation between the number of *EDN1* hotspots in a dorsal aorta section and the number near Runx1+ IAHCs (Fig.52b, $R = 0.68$, $p = 0.011$).

EDNRB cells were found in the endothelial cells, sometimes at the base of IAHCs and frequently near *EDN1* hotspots (Fig51a, c and 53b). There were also *EDNRB* cells in the mesenchyme lateral and ventral to the aorta (Fig.53b). Images of the whole tissue section revealed a stream of these cells running from the spinal ganglia and down through the spinal nerves as well as in the mesenchymal cells below and lateral to the dorsal aorta and continuing up and around the inside of the abdominal wall mesenchyme (Fig.53a). The walls of the midgut loop were also positive for *EDNRB* expression. *EDNRA* expression was less resolved to individual cells but more spread across the mesenchymal cells surrounding the aorta and, in some sections, exhibited a stronger ventral signal (Fig.53b). There was evidence also of *EDNRA* signal from the mesenchyme surrounding the vitelline artery (Fig.53a). Both receptors are expressed in the neural tube.

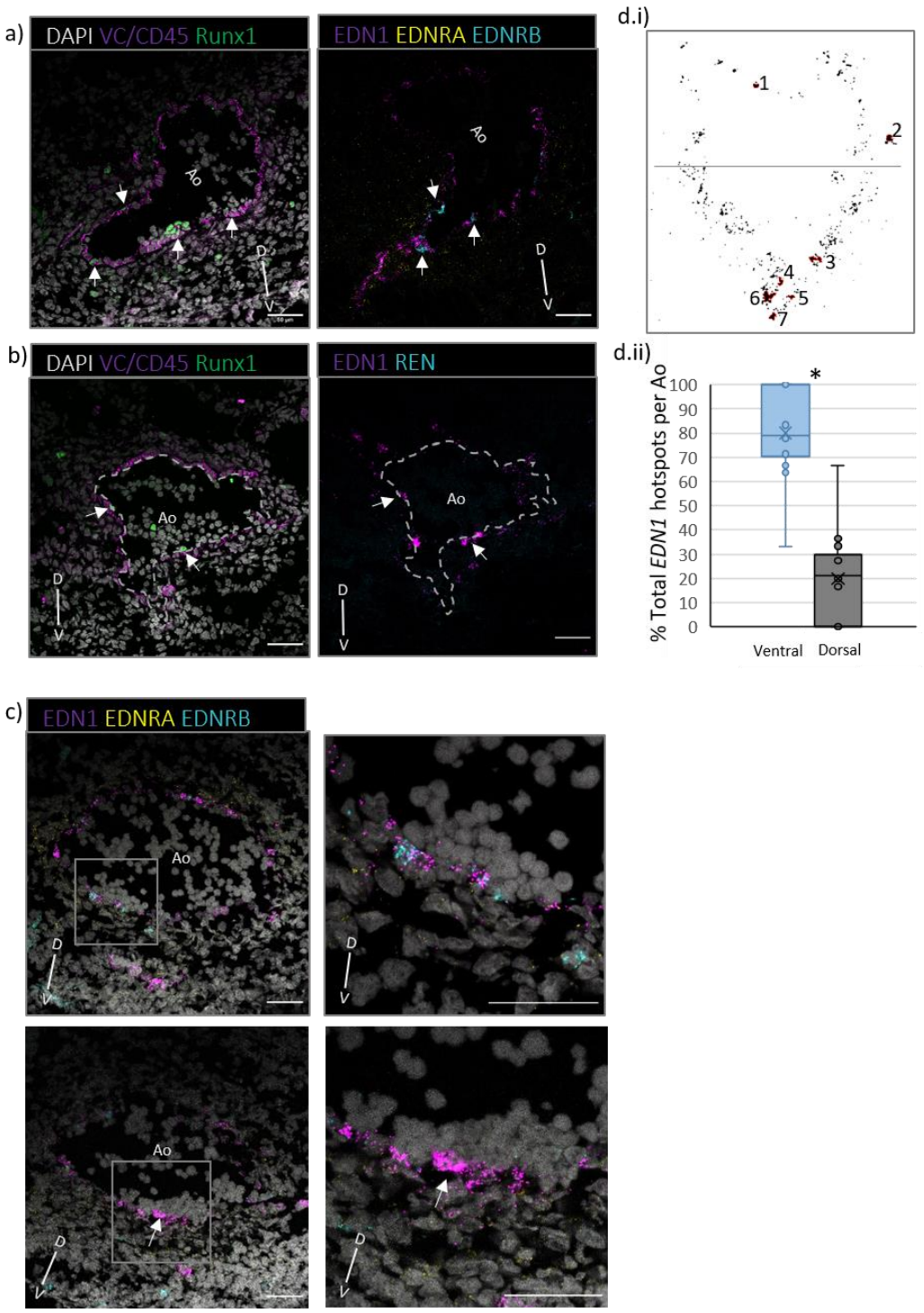


Figure 51. Visualisation of Renin and Endothelin pathway components in the AGM. All sections show the dorsal aorta (Ao) from a CS16 embryo. a-b) Protein immunostaining of Ve-Cad(VC), CD45 and Runx1 with a sister section RNAScope showing gene expression of a) *EDN1, EDNRA, EDNRB* and b) *EDN1* and *REN.* Arrows indicate Runx1+ clusters or endothelium cells and corresponding position on sister section with nearby *EDN1* hotspot (>300 pixels) or *EDNRB* expression. c) RNAScope gene expression of *EDN1, EDNRA, EDNRB.* IAHC indicated by the square and zoomed in image of the area is shown to the right. Arrow indicates *EDN1* hotspot below IAHC. di) Binary image of a section of the dorsal with the *EDN1* hotspots indicated in red. d.ii) Quantification of *EDN1* hotspots in the ventral and dorsal aorta of all sections spanning the rostral-caudal axis of the AGM (14 sections). D=Dorsal V=Ventral. Scale bar = 50µm

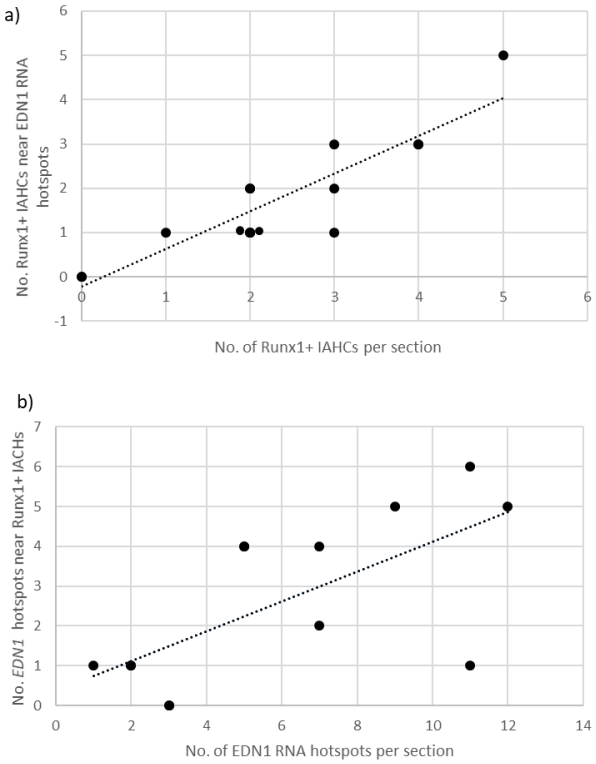


Figure 52. Correlation of position of *EDN1* hotspots with position of Runx1+ hotspots. a) The number of Runx1+ IAHCs positively correlates with the number of Runx1+ IAHCs near *EDN1 RNA* hotspots. *R* = 0.9 p=1.4e-05 b) Correspondingly, the number of *EDN1* hotspots near Run1+ IAHCs positively correlates with the number of *EDN1 RNA* hotspots. *R* = 0.68, p=0.011.

Immunohistochemical analysis was also carried out to see how the protein counterparts were expressed. Intriguingly, ET-1 (Endothelin-1 protein) was found expressed in rounded cells budding from the endothelium suggesting that it is expressed on haematopoietic progenitors (Fig.54a,d). Expression of *EDN1* was found on the VC+CD45- endothelium and not the VC+CD45+ HSPC population however (Fig.41). This raises the notion that ET-1 protein expression may be secreted and sequestered onto progenitor cells from underlying endothelial cells or that gene/protein expression may be on an earlier CD45- haematopoietic progenitor population. ET-B (Endothelin receptor B, protein) was also seen expressed on single budding cells in the ventral domain as well as expressed at the top of IAHCS (Fig.54a, b). Corresponding with RNA expression, ET-A (Endothelin receptor A, protein) was found expressed in the mesenchymal cells around the dorsal aorta (Fig54c), although no clear D-V polarisation was obvious.

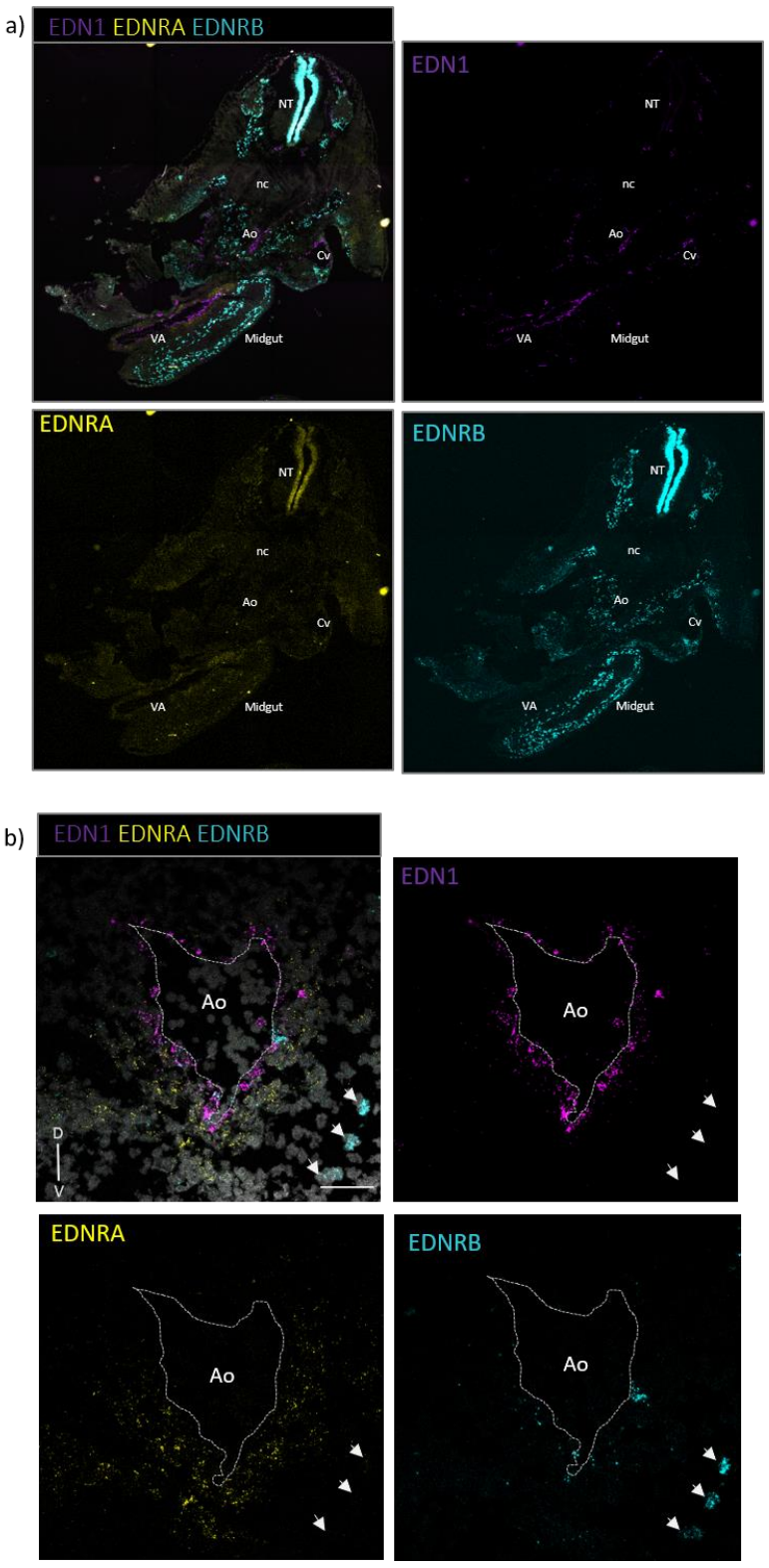


Figure 53. Visualisation of endothelin-1 and its receptors' gene expression across the embryo and dorsal aorta. All sections are from a CS16 embryo. a-b) RNAScope multiplex visualisation of *EDN1, EDNRA, EDNRB* across the whole embryo transverse section in a) and across the dorsal aorta (Ao) in b). Arrows in b) indicate single *EDNRB* expressing cells in the sub-aortic mesenchyme. D=Dorsal V=Ventral. cv= cardinal vein. VA = vitelline artery. NT = neural tube. nc = notochord. Scale bar = 50µm

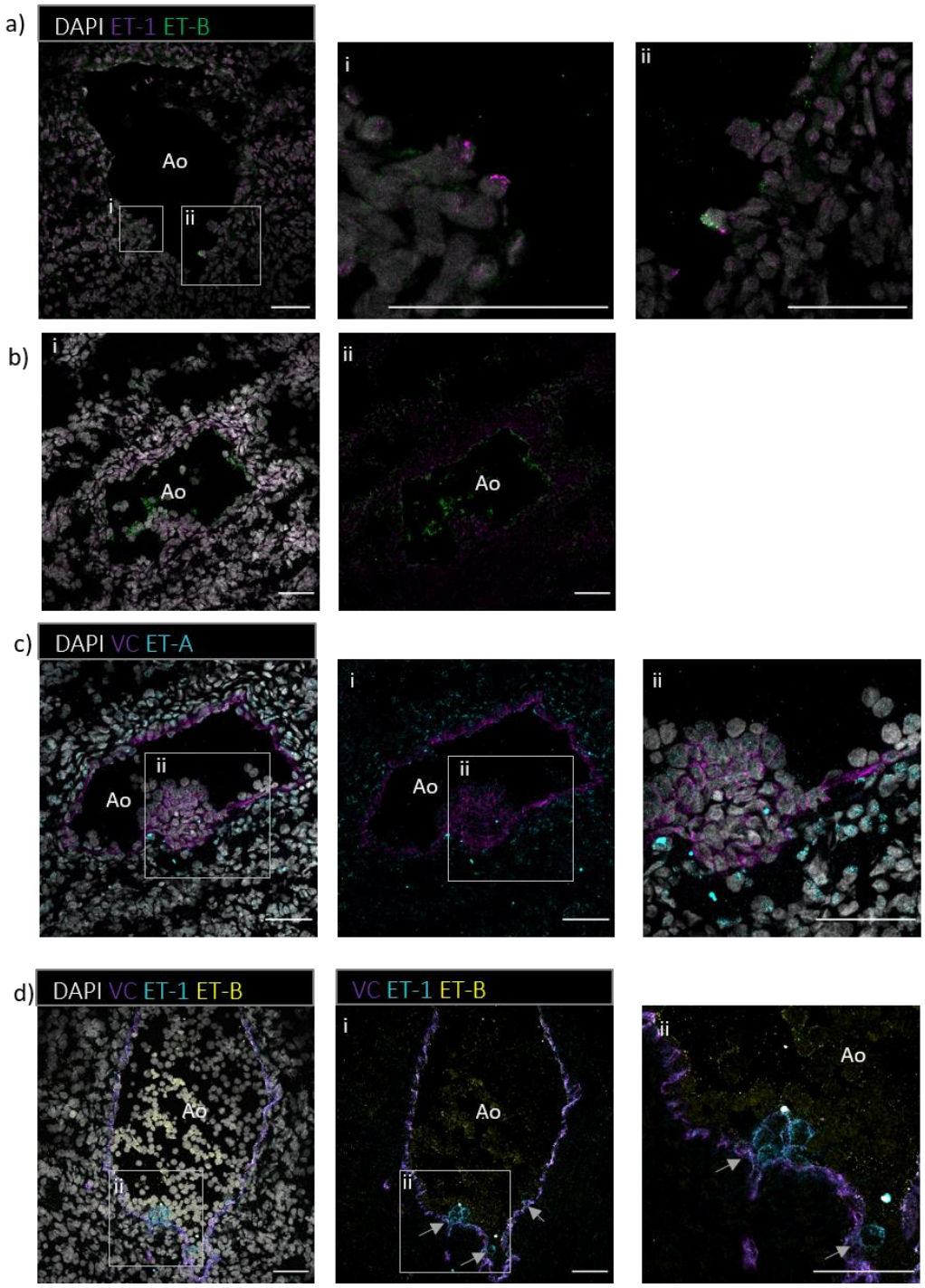


Figure 54. Visualisation of endothelin-1 and its receptors' protein expression across the embryo and dorsal aorta. All sections are transverse sections of the dorsal aorta (Ao) from CS16 embryos. a) ET-1 and ET-B expression with zoomed in images of ET-1 and ET-B budding cells (i, ii). b) ET-1 and ET-B expression with (i) and without (ii) DAPI. c) VE-Cadherin (VC) and ET-A expression with and without (ii) DAPI. The box highlights the IAHC shown in (ii). d) VC, ET-1 and ET-B expression with and without (ii) DAPI. Box highlights IAHC shown in (i). Arrows indicate emerging ET-1 expressing cells. Scale bar = 50µm

4.3.2 **REN expression**

REN expression was also characterised by RNAScope, and despite being found very highly expressed in the ventral domains of the CS17 and CS16 LCM-Seq datasets it was found expressed in just a few cells directly below the ventral endothelium (Fig.55b) These cells were also found close to *EDN1* hotspots. There was notably higher expression of *REN* along the vessels branching ventro-laterally from the dorsal aorta towards the mesonephros (Fig.55a, biii). These thin vessels had high *EDN1* expression within the VC+ endothelial cells and high *REN* expression on the cells along the outside of the VC+ cells which may be perivascular. There were also strong *EDN1* and *REN* expressing cells in the mesonephros (Fig.55a). Similarly, immunostaining with a renin (REN) antibody showed strong expression along the outside of the ventral branching vessels (Fig.55c). It is probable that the high ventral expression of *REN* from the LCM-Seq analysis resulted from including cells from the ventro-lateral branching vessels within the Mid sub-domain.

Taken together, *REN* and *EDN1* are both ventrally enriched with *EDN1* having an expression pattern more closely linked to the IAHCs suggesting a more direct influence on the processes involved in EHT. However, that is not to say the low levels of *REN* expression seen underlying the aorta aren't sufficient for an influence, and the expression from the ventro-lateral branching vessels may have an indirect effect on EHT.

4.4 **Quantification of Runx1 clusters**

The systematic staining of sections along the AGM for VC+CD45+Runx1+ cells provided an opportunity to observe where the IAHCs most frequently develop along the D-V axis. It was noticed that where the major ventral vessels, the vitelline and coeliac artery, branch off from the dorsal aorta is frequently where larger clusters are found. In sections where a ventral vessel was branching from the dorsal aorta (as in Fig.51a) the positions of IAHCs were quantified and found to come almost exclusively from within the branching vessel or at the junction with it (Fig.56). This suggests that IAHCs emerge preferentially at these junctions. Intriguingly, these junctions are also where you find strong *EDN1* hotspots (Fig.51b, Fig.53b).

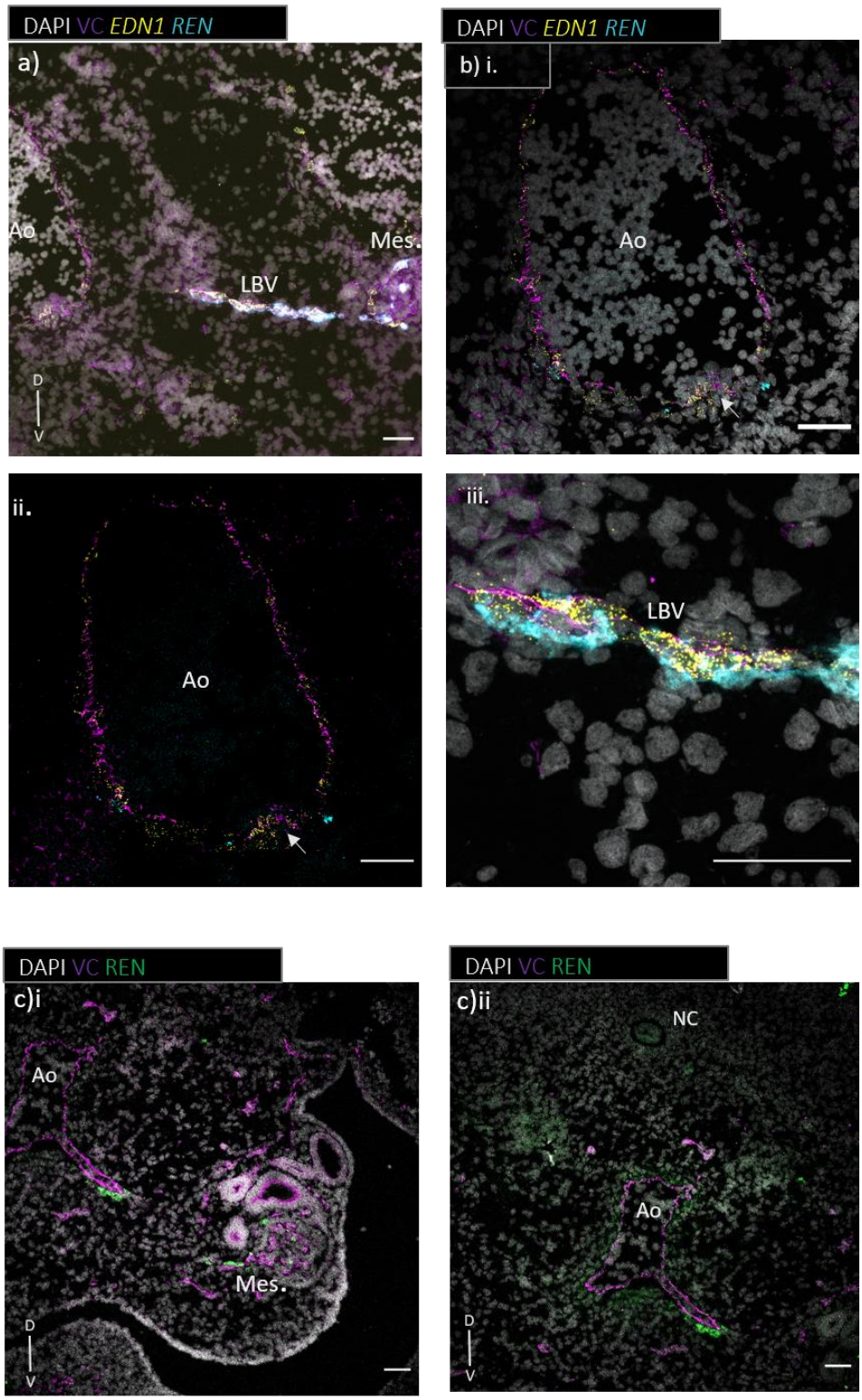


Figure 55. Visualisation of renin gene and protein expression across the AGM. All sections are transverse sections of the dorsal aorta (Ao) from CS16 embryos. a) VE-Cadherin (VC), *EDN1* and *REN* expression in the AGM showing high *EDN1/REN* expression in the ventro-lateral branching vessel (LBV) to the mesonephros (Mes.). Image taken with fluorescent microscope. b) Same section as a) but with confocal microscope. i. and ii. show dorsal aorta with (i) and without (ii) DAPI. Arrow indicates IAHC with high expression *EDN1* and spots of *REN* either side. iii. Zoomed confocal image of branching vessel in a) . c) VE-Cadherin (VC) and renin (Ren) expression . D=Dorsal V=Ventral. Scale bar = 50µm

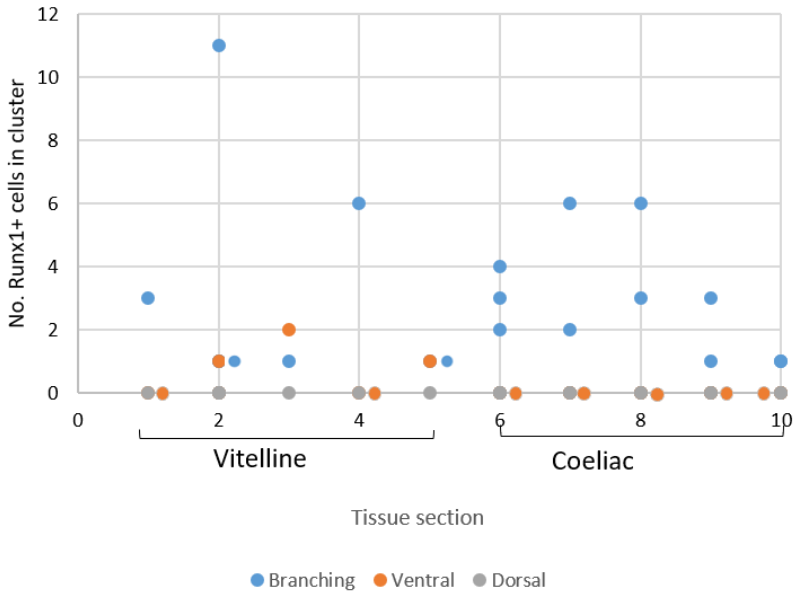


Figure 56. Quantification of Runx1+ cells per IAHC for each position in the AoV, AoD and ventral branching vessels. The number of Runx1+ cells per IAHC were counted in CS16 transverse tissue sections stained with a Runx1 antibody where the dorsal aorta branched ventrally into either the vitelline or coeliac artery. The position of each IAHC was noted; AoV (Ventral, orange), AoD (Dorsal, Grey), Branching vessel (Blue). The numbers along the X-axis denote different tissue sections. If there were no Runx1+ cells at one position the number of Runx1+ cells in IAHC = 0.

4.5 Discussion

The cells of the dorsal aorta and surrounding tissues represent a dynamic niche of multiple cell types and many signalling pathways which intersect to influence and promote the emergence of HSCs. Due to the ventrally polarised emergence of HSCs and IAHCs in the dorsal aorta as well as the demonstrated importance of polarised signalling in promoting HSC generation, it is important to define positional information when investigating signals in the niche. This was achieved with the spatial transcriptome dataset from Chapter 3 which elucidated polarised gene and pathways expression both proximal and distal to the dorsal aorta, highlighting factors with a potential role in influencing HSC emergence such as the ventrally enriched *REN* and *EDN1* or the TGF-β and TNFα via NFκB pathways. However, the spatial transcriptome data represents just one dimension of the niche; relative position to the dorsal aorta and site of HSC emergence. The RNA-Seq analyses in this chapter retained some spatial identity by sub-dissecting the AoV and AoD, but also explored cell population specific transcriptomes. The combined spatial transcriptome and cell population analyses builds up a more complete picture of the niche and enables modelling population interactions for experimental analysis.

4.5.1 Clues to the processes behind endothelial-to-haematopoietic transition

The bulk-seq analysis of haemato-endothelial populations in the dorsal aorta was used to define an HSPC signature including several TFs. These include *MYB*, which has been shown to generate short-

term myeloid-erythroid progenitors (MEPs), and *GFI1* which has been used to reprogram murine endothelial cells into HSCs (Yokomizo *et al.*, 2019; Doulatov *et al.*, 2013; Lis *et al.*, 2017). This suggests that this VC+CD45+ population contains both HSCs and more committed progenitors. Further HSPC signature TFs could be tested for their ability to reprogram PSCs to HSCs such as the highly HSPC specific *SETBP1* and *PRDM16*. Notably, the receptors upregulated on the HSPC population vs. the endothelial population were similarly expressed in the more mature VC-CD45+ population. This may explain in part why it has been a challenge to refine the markers that define the HSPC population.

The HSPC signature provides a snapshot of the gene expression of HSCs and progenitors but does not necessarily reveal the signals and processes required to become those cells. In order to identify the signals that initiate HSC specification, it is useful to identify the endothelial fraction destined to become haematopoietic, and for this the single cell analysis was useful. It revealed that the AoV endothelium contains heterogenous endothelial cells with distinct expression signatures. A central network of 4 clusters (clusters 1,2,5 and 7) of endothelial cells which had similar expression signatures (Fig.46b) could be considered the default endothelial cell type with more specialised endothelial clusters branching off: 1) *FGF23* expressing cells with enrichment of Hypoxia and mTORC1 signalling pathway (cl.8). 2) *GJA5* expressing cells with an enrichment of ECM associated proteins and TNFα via NFκB and Epithelial-Mesenchymal Transition pathways (cl.0).

Within the central group of endothelial cell clusters (cl. 1,2,5 and 7) one cluster, cl.5, contained a fraction of double positive VC+CD45+ cells. The lineage inference analysis showed lineage links of cl.5 with the haematopoietic cl.6 but also cl.0 *GJA5+* endothelial cells. The cl.5 endothelial cells had no discerning markers from the other central network cells other than proliferation associated genes which were markedly reduced in the VC+CD45+ population. There is a possibility that HSC specification genes were not captured in this dataset or that the cells expressing these genes were too few to cluster separately. The data do suggest however that an increase in proliferation precedes differentiation towards *GJA5* endothelial cells or haematopoietic cell fate. This parallels an increase in the proliferative status of pre-HSCs in the mouse which reduces as they acquire a adult-type HSC state (Batsivari *et al.*, 2017). Further experiments to investigate the effects of proliferation status on haematopoietic output would be useful. Furthermore, the lack of defining gene markers on this sub-population of 'haematogenic' endothelial cells could indicate that a different layer of regulation is involved in specifying endothelial cells to become haematopoietic. For example, gene expression may be modified post-transcriptionally by non-coding RNAs such as micro-RNAs which are not detected by the RNA-Seq methods in this project.

Another insight from these data is the enrichment within the *GJA5+* endothelial cells (cl.0) of the Epithelial-Mesenchymal Transition (EMT) gene set. Cells undergoing EMT dissociate from a structured epithelium with high cell-cell adhesion and apico-basal polarity (Campbell and Casanova, 2016). This

requires changes in morphology and therefore actin-myosin skeleton organisation, a break-down of cell adhesion and polarity, and loose, dynamic attachments with ECM facilitating motility and migration. The process has been likened to EHT during which a haematogenic endothelial cell must simultaneously round its morphology whilst dissociating from the structured endothelium, become more motile and bud into the lumen (Kissa and Herbomel, 2010). It is reasonable to assume that EHT and EMT share mechanisms to facilitate this cell movement. It is possible that this sub-population of endothelial cells (cl.0), even if they are not necessarily the cells undergoing EHT they may facilitate this process. It would be useful to see if markers from cl.0 correspond to the localisation of IAHCs. Indeed, high expression of *EDN1*, correlates highly with the presence of IAHCs, although a more specific marker like *GJA5* may be more revealing. Taken together, the single cell data suggest a model whereby endothelial cells in cl.5 increase their proliferation status and differentiate towards either a haematopoietic fate or to specialised endothelial cells which in turn may enhance EHT through signalling or modulation of the ECM interactions.

4.5.2 **Lessons from the datasets**

This chapter investigated the transcriptomes of individual cell types of the haemato-endothelial compartments of the dorsal aorta and niche cell sub-types both through bulk population and single cell RNA-Seq. With these additional data, signalling in the niche could be resolved to distinct cell populations.

i. TNFα via NFκB

The TNFα via NFκB pathway, which was ventrally enriched in the cells of the vessel wall and sub-aortic mesenchyme was also enriched in the haematopoietic and HSPC (VC+CD45+) population compared with the VC+CD45- endothelial populations. However, further resolution of the endothelial populations by single cell revealed that this signalling pathway is enriched on endothelial sub-types (clusters 0 and 2), including upregulation of pathway target genes. There were also TNF receptors on both the endothelial (VC+CD45-) and HSPC (VC+CD45+) populations. This, along with the fact that *TNF* itself is expressed in the haematopoietic cells, suggests that TNFα via NFκB signalling initiated in haematopoietic cells stimulates a response in the endothelial, HSPC and other niche cells through different receptors. Indeed, in the zebrafish it has been shown that TNFα provided by neutrophils stimulates EHT via Notch signalling, specifically the Notch ligand Jag1 (Espin-Palazon *et al.*, 2014) (Fig.57a). Intriguingly, in this single cell dataset, cl.0 which had enrichment of TNFα via NFκB pathway and highest endothelial expression of *TNFRSF1B* (TNF receptor 2) also had enrichment of Notch targets (*HEY2)* and ligands including *JAG1.* In the mouse, yolk-sac derived macrophages in the sub-aortic mesenchyme, express inflammatory genes including *Tnf,* migrate to the dorsal aorta, intravasate the vessel wall and interact with emerging IAHCs (Mariani *et al.*, 2019) (Fig.57b). The authors reported that this process was critical for HSC emergence. Macrophages have been shown to accumulate below

the endothelium in the ventral part of the dorsal aorta in humans (Travnickova *et al.*, 2015). It is tempting to postulate that a similar process occurs in the human embryo: proinflammatory macrophages or neutrophils produce TNFα which acts on a specific endothelial sub-population (cl.0) via TNFR2, stimulating Notch signalling through production of Notch ligands and thus promoting EHT.

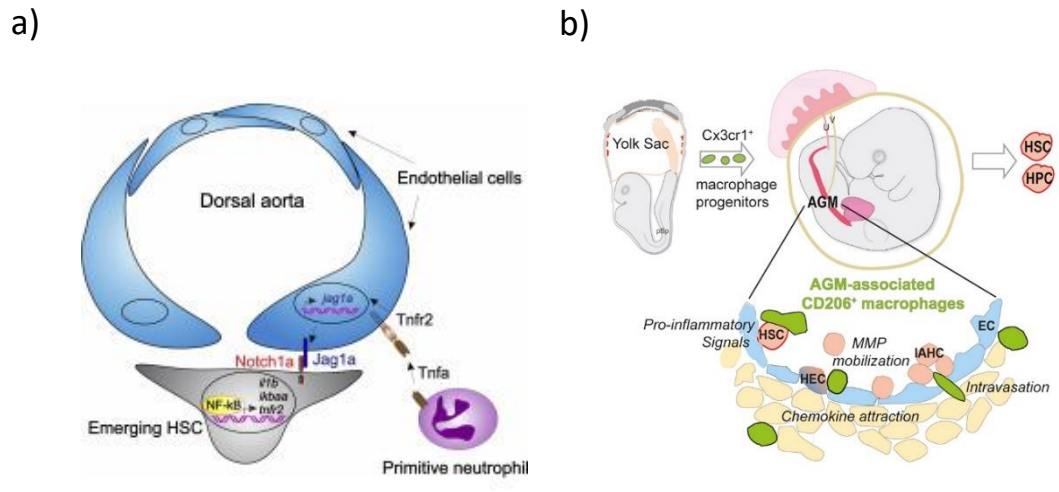


Figure 57 TNFα signalling in the zebrafish and mouse models. a) Taken from R. Espin-Palazon et al., 2014 (Espin-Palazon *et al.*, 2014). Primitive neutrophils provide Tnfa signal which is received onto endothelial cells of the dorsal aorta via the receptor Tnfr2. This activates Notch signalling stimulating endothelial-to-haematopoietic transition. b) Taken from S. A. Mariani et al., 2019 (Mariani *et al.*, 2019). Primitive macrophages which express an inflammatory signature including Tnf migrate to the endothelium of the dorsal aorta, intravasate and associate with the emerging IAHCs – a process that is critical for HSC emergence.

ii. *mTorC1 signalling*

The mTORC1 signalling pathway was enriched in the LCM-Seq ventral Mid vs dorsal Mid as well as in the Inner and Mid layers vs Outer. The single cell analysis from this chapter resolved this signalling pathway to an endothelial cluster (8). As discussed in the introduction, a previously published single cell analysis of E11 endothelial, pre-HSCs discovered that pre-HSC I has an enrichment of mTOR pathway and disruption of the mTORC2 pathway inhibited HSC emergence in the AGM (Zhou *et al.*, 2016). It is unlikely that cl.8 represents pre-HSCs as there is no strong lineage relationship with cl.5 which contains HSPCs or haematopoietic cl.6. However, mTORC1 pathway is also enriched at lower levels on cl.5. Notably, hypoxia pathway is also enriched in cl.8 which acts upstream and regulates mTORC1 signalling. This suggests that this sub-population may sense hypoxia and mediate downstream mTORC1 signalling on the pre-HSC population.

iii. Osteopontin

Osteopontin is a secreted phosphorylated glycoprotein that, in the bone marrow niche, mediates HSC migration and proliferation (Guidi *et al.*, 2017; Nilsson *et al.*, 2005). However, its role in the embryonic haematopoietic niche is not characterised. In these datasets it was upregulated in the ventral sub-aortic mesenchyme (LCM-Seq Mid) but also on the VC+CD45+ HSPC and VC-CD45+ haematopoietic populations indicating that its expression in the ventral mesenchyme may derive from CD45+ haematopoietic cells such as macrophages. Indeed, osteopontin expression on activated macrophages has previously been described (Zohar *et al.*, 2000). Further immunostaining studies would confirm this. Intriguingly, osteopontin directly binds the two top enriched receptors of the HSPC population (compared with endothelial), CD44 and ITGA4. However, a subset of endothelial cells, cl.8, express the gene for acid phosphatase 5 which regulate osteopontin activity (O'Leary *et al.*, 2016). This expression on a small population of endothelial cells suggests localised regulation of this pathway. Furthermore, endothelin stimulates osteopontin expression (Berglund *et al.*, 2016; Shioide and Noda, 1993). It would be interesting to explore the effects of osteopontin on HSC emergence.

iv. Plasminogen activators

The gene PLAU encoding the secreted factor Plasminogen Activator, Urokinase, was found to be ventrally enriched in the dorsal aorta vessel wall and has a large number of known interactions with receptors on the ventral HSPC population. Further experiments are required to define the cell origins of *PLAU* in the human AGM as it was not detected in the single cell dataset. However, *PLAT,* another ventrally enriched plasminogen activator was detected specifically on the *EPCAM*+ cl.9 which may be neuroendocrine cells due to their expression of *GATA3* and parathyroid hormone (*PTH).* Notably, in the bone marrow plasminogen activation releases Kit ligand (SCF) from stromal cells resulting in increased haematopoietic progenitor proliferation and differentiation (Heissig *et al.*, 2007). SCF has been shown to stimulate HSC production in the mouse embryo (Rybtsov *et al.*, 2014). It would be interesting to see if PLAU acts in a similar way to make SCF more available in the human embryo dorsal aorta.

v. Erythromyeloid progenitors

The single cell dataset indicated a cluster (cl.10) of CD34+VC+ cells which express embryonic and foetal haemoglobins and had an RNA velocity towards the central endothelial network indicating that they contribute towards this pool of endothelial cells. Furthermore, the dorsal endothelial (VC+CD45-) cells had enrichment of embryonic and foetal hemoglobins (*HBE1*, *HBZ, HBG1, HBG2)* compared to the ventral domain adding further evidence of a link between aortic endothelium and primitive erythroid progenitors. Indeed, it has been reported that embryonic erythro-myeloid progenitors can contribute to the endothelium of blood vessels in the embryo proper (Plein *et al.*, 2018). If a contribution of

embryonic haemoglobins to the aortic endothelium is true then a re-evaluation of the EHT paradigm is required. However, it cannot be absolutely ruled out that this is not an artefact of RNA Velocity as it is unclear whether all embryonic cell types have equivalent splice kinetics which may affect the direction of velocity. This relationship needs to be demonstrated functionally and with further single cell analyses before any real conclusions can be drawn.

It must be noted that the single cell data were just one replicate and that further single cell analyses would strengthen the conclusions drawn from this analysis. The 'contamination' by non-CD34 expressing cells in the single-cell dataset provided the possibility of an exploration of the niche cells which was of great use. However, it is highly likely that not all niche cell sub-types were captured as expression of certain genes found highly expressed in the ventral mesenchyme, such as *REN* were not found in the dataset indicating that these cells were missing. It is unlikely that this is the result of drop-outs, where the gene is below the threshold of expression detected by the datasets, as expression levels were high in the LCM-Seq dataset. This dataset also did not let us explore the HSPC population in much resolution as the cell numbers captured were small. An alternative single-cell technique such as Smart-Seq2 may be more appropriate for this as it allows for deeper sequencing of small sorted populations.

The analyses described in this discussion demonstrates how application of what is known in model organisms can be consolidated with the datasets in this work to uncover probable analogous mechanisms of HSC specification such as for TNFα signalling. These can then be a starting point for experiments to functionally validate the proposed mechanisms. The combined datasets are also a source of previously unexplored potential players such as the plasminogen activators, osteopontin and novel HSPC specific TFs, to name a few. The next section describes how these datasets can be used to functionally explore novel players which is the focus of the rest of this thesis.

4.5.3 **Using the dataset as a starting point for functional validation experiments**

The combined spatial transcriptome, bulk-seq and single cell analyses provide a detailed, multi-dimensional description of the developmental haematopoietic niche. It has revealed many genes and pathways with a potential influence on HSC emergence and which may serve as the foundation of future functional assays. In this work, *REN* and *EDN1* have been the focus due to their ventral enrichment and relevant downstream pathways (see Chapter 3 discussion 3.4.5). Resolution of their central pathway components found only *ACE* marks the HSPC population itself which is in line with previous reports that it is an HSC marker in the human embryo (Jokubaitis *et al.*, 2008; Sinka *et al.*, 2012). By immunostainings and in situ hybridisation, *REN* itself was found most strongly expressed surrounding the endothelial cells of ventro-lateral branching vessels although weaker expression was seen localised in spots beneath the ventral aortic endothelium. It will be interesting to see if these spots correspond to endothelial expression of the angiotensin receptor *AGTR1* which was expressed

exclusively on the *GJA5* endothelial subset (cl.0) in the single cell analysis. This is the same cluster with the highest and most consistent expression of *EDN1*, and its activating enzyme *ECE1.* It would also be interesting to see whether the *EDN1* hotspots seen on the RNAScope analysis correspond with *GJA5* expression.

EDN1 was highly upregulated in the endothelial cells compared to the progenitors, and although expression of *EDN1* may not contribute towards HSC identity itself, the high correlation of *EDN1* hotspots with sites of Runx1+ IAHCs suggests that it may be an important niche signal. Furthermore, cl.0 contains Notch ligands *JAG1, DLL4,* and downstream TF *HEY2* adding further evidence that these cells are important niche cells. Notably, ET-1 (endothelin-1 protein) treatment of giant cell tumour stromal cells increases expression of Jagged1, Hey2 and Mmp2 (also cl.0 enriched), increasing cell growth and migration (Yuan *et al.*, 2019). *EDNRB* expression was both on the endothelial cells themselves and on the mature haematopoietic cells as shown on the bulk-seq, single cell, RNAScope and IHC analyses. ET-B (endothelin receptor B, protein) was seen expressed on the top of the IAHCs, yet *EDN1* is expressed in the endothelium. This leaves a gap of VC+CD45+ progenitors between them raising the question of whether ET-1 is secreted across them. The immunostainings indicate that although *EDN1* is expressed on the endothelial cells, the protein can be seen expressed on rounded cells at the bottom of IAHCs suggesting that the progenitors express ET-1 at the protein level.

EDNRB expression in the endothelium was seen on individual cells interspersed throughout the endothelium. Further co-staining experiments of the *EDNRB* probe with IAHC marking antibodies may help define whether selective *EDNRB* expression is related to HSPC emergence. The *EDNRB* also had a pattern of expression on cells streaming down from the spinal ganglia to the ventral mesenchyme below the dorsal aorta consistent with the ventral migration pattern of the neural crest, Notably, in the mouse, neural crest cells fated to become melanocytes and enteric neuronal cells express *Ednrb* and at E9.5 are found in ventro-lateral positions adjacent to the bifurcated dorsal aorta (Lee, Levorse and Shin, 2003). In zebrafish neural crest cells migrate towards the dorsal aorta making contact with it at the onset of HSC emergence and perturbation of this contact leads to loss of HSCs (Damm and Clements, 2017). Previous studies suggest that sympathetic ganglia derived from the neural crest lodging beneath the dorsal aorta influence HSC emergence (Fitch *et al.*, 2012). This study indicates that these effects may be mediated by endothelin signalling.

4.5.4 **Concluding remarks**

Combining the spatial transcriptome dataset with single-cell and sorted bulk population sequencing as well as fluorescent microscopy has resolved polarised signalling pathways to discrete populations facilitating modelling of potential mechanisms behind HSC emergence. However, the data in this thesis have thus far been descriptive and the proposed mechanisms theoretical. As proof of the utility of this

dataset, the next chapter aims to define a functional role for endothelins and renin in promoting HSC emergence in the AGM.

5. Functional analysis of the role of Renin and Endothelin in promoting HSC generation

In the preceding chapters, renin and endothelin-1 were identified as candidates with a potential role in promoting HSC generation. In this chapter the mouse floating reaggregate culture system is used to assess whether these factors can increase HSC numbers derived from the AGM and how they influence haematopoietic output.

5.1 **Renin and endothelin pathway genes are enriched in the ventral domain of the mouse AGM during HSC development**

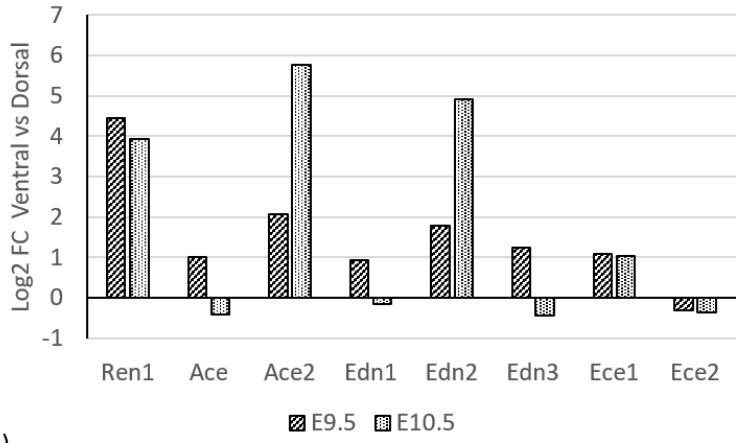
A recent publication from this lab investigated D-V transcriptome differences in the mouse AGM at the time of development of HSCs (E10.5) and their precursors pre-HSCs (E9.5)(McGarvey *et al.*, 2017). Genome-wide RNA sequencing was carried out on the sub-dissected dorsal and ventral domains of AGMs pooled from 15-34 embryos in 3 experiments, providing robust spatial transcriptome data. Using this dataset, the polarisation of central renin and endothelin pathways' genes could be elucidated (Fig.58). This included genes encoding renin and endothelin isoforms and the angiotensin and endothelin converting enzymes. Every gene was ventrally enriched at either E9.5, E10.5 or both except for *Ece2* which was slightly dorsally enriched. *Ren1* has some of the lowest expression levels of all the genes but exhibited consistently high differential expression in the ventral vs dorsal domain. *Ace2* had high log2 fold change (FC) enrichment in the ventral domain at E9.5 and E10.5 whereas *Ace* had a more modest log2 FC enrichment at E9.5 only, suggesting that in mouse *Ace2* and not *Ace* may be the RAS effector influencing haematopoiesis. Both *Edn1* and *Edn3* showed ventral enrichment at E9.5 only whereas *Edn2,* although expressed at much lower levels, had a much higher ventral enrichment at both E9.5 and E10.5. *Ece1* was the most highly expressed gene taken for this dataset and had a roughly two-fold ventral enrichment at E9.5 and E10.5.

Transverse sections of E10.5 embryos were immunostained using antibodies for CD31 and cKit, to visualise IAHCs, and ET-1 (Endothelin-1). Intriguingly, within the dorsal aorta ET-1 was found to be expressed most highly on the cKit+ IAHCs on the cell membrane (Fig.59). Expression was often localised to the membrane exposed to the lumen of the dorsal aorta. There was also lower expression throughout the sub-aortic mesenchyme.

Taken together this shows that both renin and endothelin pathway gene expression is polarised in the mouse AGM indicating analogous roles in the AGM during HSC development. Furthermore, ET-1 expression on the IAHCs indicates a signalling role on the HSPCs. This opens up the mouse model as a system for exploring the functional roles of these pathways in HSC development providing a mechanism paralleling that in the human embryo. As *Ren1* and *Edn1* are homologues of the genes

from the human dataset these were used in further functional validation assays along with the addition of *Edn2* due to its high ventral enrichment.

a)



b)

Gene	E9.5 Ventral	E9.5 Dorsal	E10.5 Ventral	E10.5 Dorsal
Ren1	0.195	0.00896	0.381	0.0248
Ace	2.16	1.07	4.65	6.18
Ace2	0.218	0.0519	2.75	0.0507
Edn1	9.58	4.98	18.6	20.7
Edn2	0.0308	0.00896	1.02	0.0336
Edn3	22.1	9.36	13.9	19
Ece1	206	96.6	225	109
Ece2	18.6	23.1	15.5	19.9

Figure 58 : Expression of renin and endothelin pathway members in spatial transcriptome data from A. McGarvey et al., 2017. a) Log2 fold change of the AoV vs AoD at E9.5 and E10.5. b) Fragments Per Kilobase of transcript per Million mapped reads (FPKM) normalised values for each gene in AoV and AoD at E9.5 and E10.5

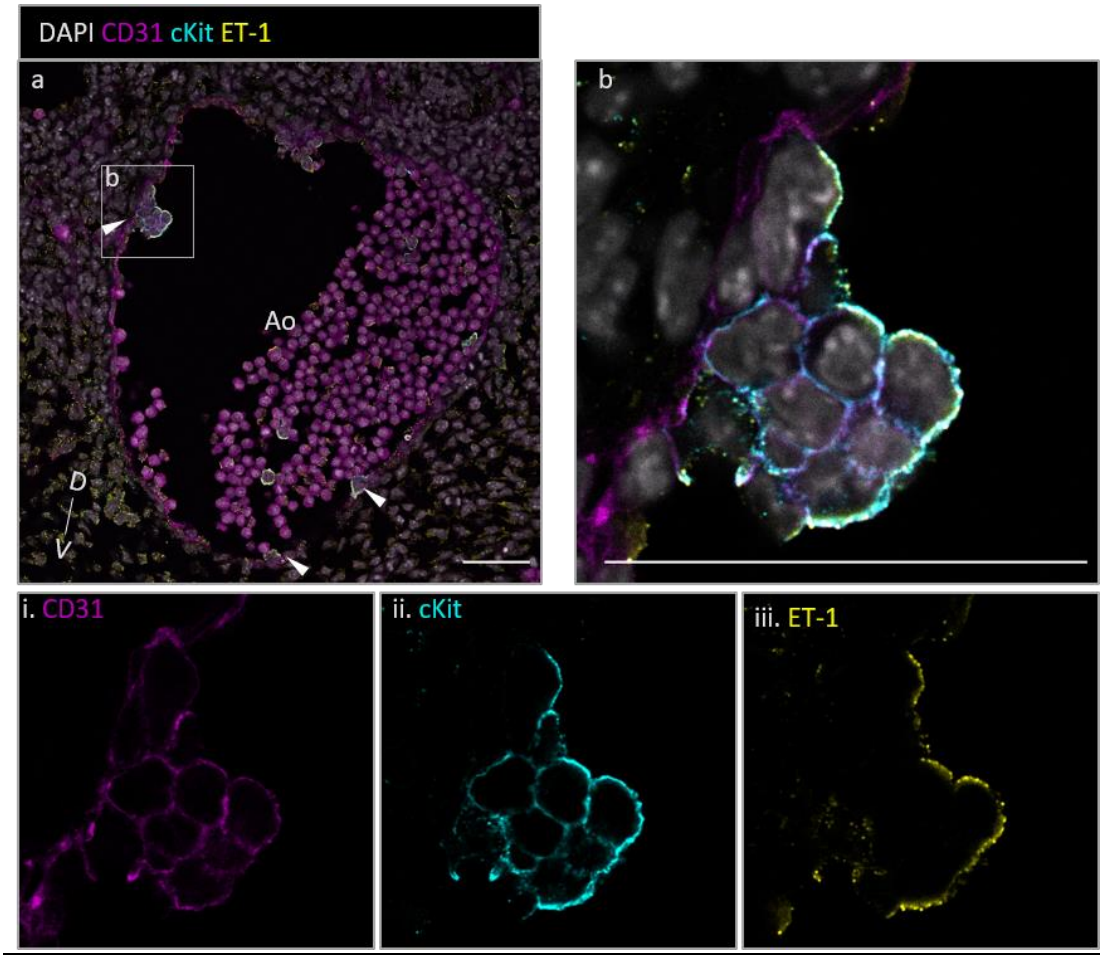


Figure 59 : ET-1 expression in the mouse dorsal aorta. (Immunostaining done by H. Felchle). E10.5 transverse section showing the dorsal aorta (Ao) immunostained with CD31, cKit and ET-1. Arrowheads indicate cKit+ET-1+ IAHCs or single rounded cells attached to endothelium. The box indicates a lateral IAHC which is shown at higher magnification in b) i-iii show individual channels for image b. D = Dorsal. V = Ventral. Scale bar = 50µm.

5.2 Investigation into the functional role of renin, ET-1 and ET-2 in promoting HSC emergence

In order to assess whether any of the three secreted proteins are able to enhance HSC emergence E9.5 caudal tissues and E10.5 AGMs were cultured for 7 days as cell reaggregates or explants respectively, on membranes floating on media with addition of one of the 3 factors at either 10ng/µl or 100ng/µl (Fig.60). At the end of the culture period the effect of the factors on haematopoietic output was assessed through haematopoietic colony (CFU-C) assays and immunophenotypic analysis of the haematopoietic stem/progenitor and mature cell compartments. E9.5 reaggregates were also transplanted into irradiated mice in a repopulation assay to assess the effect on HSC production.

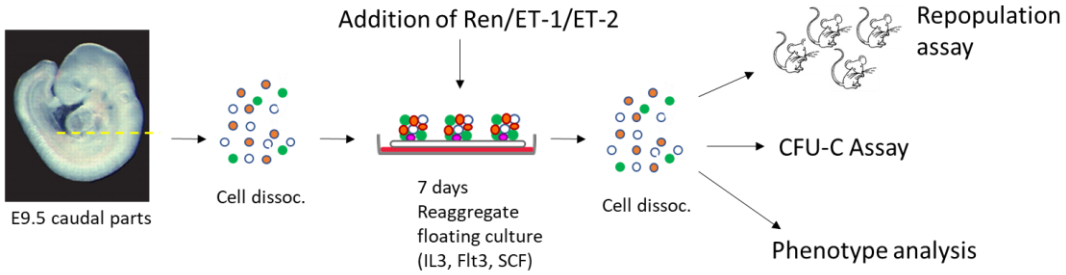


Figure 60. Experiment layout for E9.5 floating reaggregate cultures and assays to assess the effects of Renin, ET-1 and ET-2 on developmental haematopoiesis. E9.5 caudal parts were dissociated into single cells and reaggreged (1 embryo equivalent / aggregate). The reaggregates were cultured on floating membrane cultures supplemented with IL3, Flt3, SCF and addition of one of the test factors. After 7 days cells were dissociated and analysed by repopulation assay, CFU-C assay and immunophenotypic analysis.

5.2.1 *In vitro* haematopoietic colony forming-cell assay indicates a trend towards more immature colony types after culture with renin and endothelins.

E9.5 caudal tissue were dissociated into single cells and cultured for 7 days as reaggregates on a floating membrane system with addition of haematopoietic cytokines SCF, IL3 and Flt3 (SIF) as well as one of the factors of interest, renin, ET-1 and ET-2 at 2 concentrations each; 10ng/µl and 100ng/µl (Fig.60). Notably, CFU-C total numbers decreased with addition of any of the factors (Fig.61ai). A look at the numbers of different colony types showed that CFU-Mac which derive from a more mature progenitor committed to macrophage production, are the main cause for the reduction in total CFU-C numbers after culture with renin, ET-1 or ET-2 which is significant for ET-1 100 (Mann-Whitney U test, $p < 0.05$) (Fig.61aii). However, there is a trend towards production of higher numbers of more immature CFU-Cs with greater differentiation potential: colony-forming unit-granulocyte, erythroid, macrophage, megakaryocytes (CFU-GEMM), for each of the added proteins, significantly for the ET-2 100 (Fig.61aiv, Mann-Whitney U test, $p < 0.05$). Colony-forming unit-granulocyte, macrophage (CFU-GM) numbers were unaffected in each condition (Fig.61aiii).

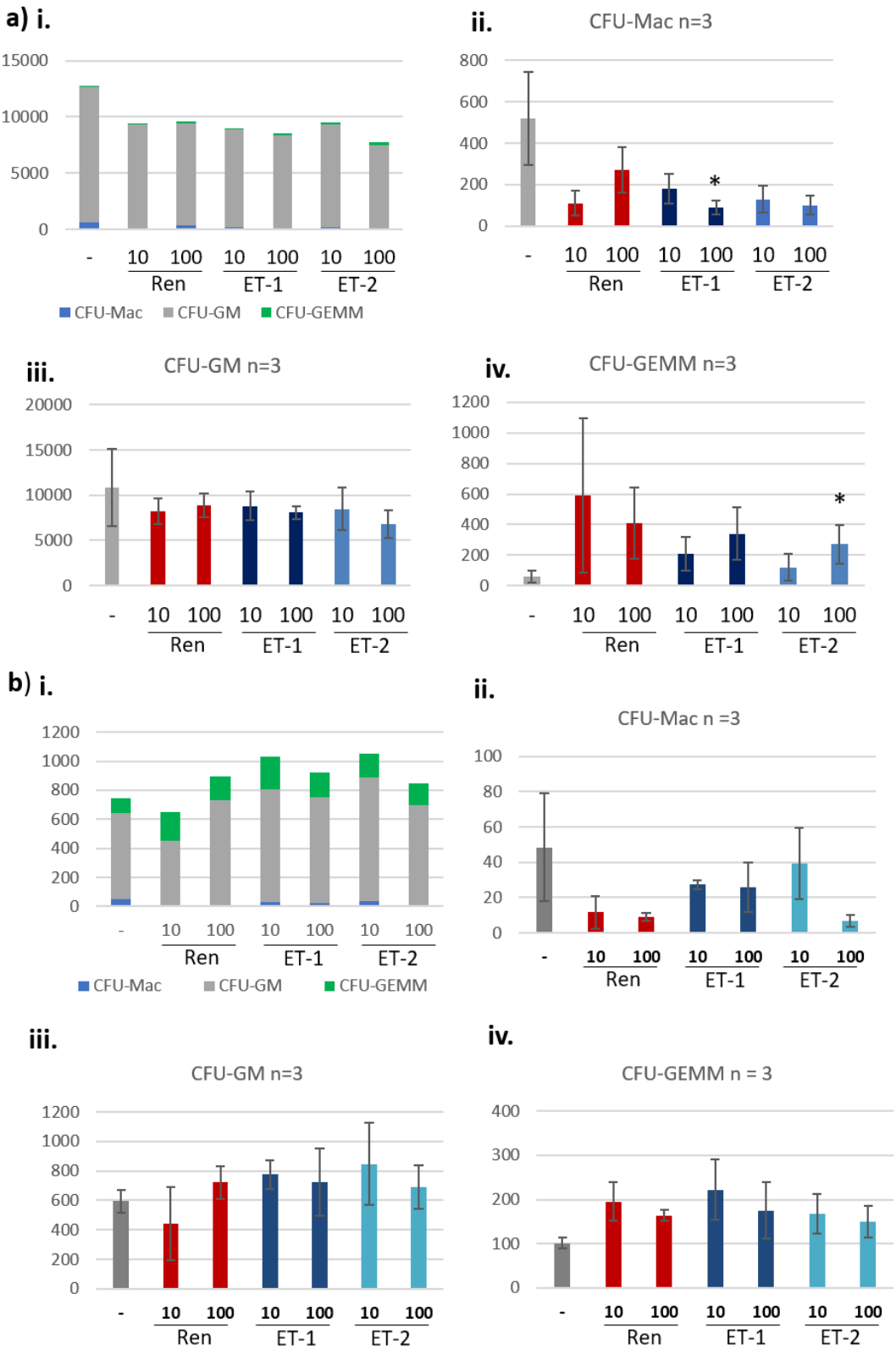


Figure 61 CFU-C numbers for a) E9.5 caudal parts (n=4) and b) E10.5 AGM cultures (n=3). i: Total CFU-Cs. ii: CFU-Mac. iii: CFU-GM iv: CFU-GEMM. Error bars indicate standard deviation around the mean. CFU-C numbers that are significantly different from the control are indicated with an asterisk (Mann-Whitney U test, $p<0.05$).

E10.5 AGM explants were also cultured for 7 days with addition of one of Ren, ET-1 or ET-2 at 10ng/µl and 100ng/µl but this time without FCS and SIF cytokines to remove potential confounding effects of these factors. At this stage there was again a trend towards higher numbers of CFU-GEMMs but also a higher numbers of CFU-GMs for each test condition with the exception of Ren 10ng/µl CFU-GM numbers (Fig.61bi,iii,iv). The increase in both these types of early progenitor colony resulted in a net increase of overall colony numbers for all test conditions except Ren 10ng/µl and this was statistically significant for Ren 100ng/µl (paired t-test, $p<0.001$). Taken together, the data highlight a potential drive of production of multipotent progenitors over more mature progenitor types when renin, ET-1 or ET-2 are added to the haematopoietic cultures of both E9.5 and E10.5 AGM cells.

5.2.2 **Phenotypic characterisation of AGM cell populations cultured with Ren, ET-1 and ET-2**

After 7 days of E9.5 reaggregate culture with Ren, ET-1 or ET-2 the cell phenotypes of the reaggregates were analysed with two panels of antibodies to investigate A) HSC/HSPC and vessel wall cell compartments (Fig.62) and B) the haematopoietic lineage compartments (Fig.64). In Panel A, the HSC_HSPC compartment was defined as VC+Sca1+Kit1+CD44+CD41^{lo} CD45+ (Fig.62a-c). From the VC+CD45-CD41-CD146- endothelial compartment, CD44 was used to define arterial vs. venous endothelium (Fig.62g). Perivascular and epithelial populations were also analysed using CD146 and EpCAM from the VC- population respectively (Fig62d-e).

For each of the test culture conditions there is an increase in the proportion of cells in the SK+ and VC-CD45+ haematopoietic cell populations compared to the control suggesting that, although not statistically significant, all 3 factors may have a positive influence on haematopoiesis in the cultures (Fig.63). However, there is no increase in the phenotypically defined HSC_HSPC population. Although modest, addition of both endothelins to culture increases the proportion of the aortic and venous endothelium compared to control. Conversely the epithelial compartment appears reduced in these conditions suggesting a drive for endothelial over epithelial production. No difference was seen in the perivascular compartment.

After 7 days of E10.5 explant culture there are not many appreciable differences in the cell populations analysed (Fig.64). However, it is of note that ET-1 10 has the highest average SK+, HSC_HSPC and VC-CD45+ haematopoietic populations and the highest number of CFU-GEMMs in the haematopoietic colony assay (Fig.61). This suggests that ET-1 has a positive influence on haematopoietic progenitor production. However, the higher concentration shows reduced levels of VC+, SK+, HSC_HSPCs and VC-CD45+ populations indicating that the effect of ET-1 is dose dependant at E10.5.

Panel B defines mature haematopoietic lineages through a series of markers to determine whether addition of any of the factors to culture skews haematopoietic differentiation towards one lineage or

another (Fig.65a). The populations analysed are defined as Monocytes and early macrophage progenitors(Mo, Gr1-CD11b+), Gr1+CD11b- granulocytes, macrophages (Mac, Gr1+CD11b+), B220-CD3+ lymphoid cells, B cells (B220+CD3-), pro-B cells (B220+CD3-CD45+CD5+), NK Cells (CD335+CD5-), unknown population B220-CD5+, erythrocytes (All-Ter119+) and dendritic cells (DC, All-CD11c+)(Fig65a(a-d)).

For E9.5 cultures an increase in the proportion of B cells within the total live cell population is seen in Ren and ET-2 higher concentrations and both concentrations of ET-1 (Fig.65b). This is accompanied by the appearance of a population positive for both the B cell and T cell markers (B220+CD3+). Such double positive cells have been found in the foetal liver defining a lymphoid progenitor cell with both T cell and B cell potential (Sagara *et al.*, 1997). This population is thus denoted as a lymphoid progenitor (Fig.66). The percentage increase in B cell and lymphoid progenitors in the Ren 100, ET-1 10 and ET-1 100 cultures (1.37, 0.9 and 1.66% respectively) are accompanied by a similar reduction in proportions of the monocyte and macrophage populations (-2.06, -1.68 and -1.48%) respectively (Fig.65b). This suggests that these factors may skew differentiation towards lymphoid rather than myeloid lineages in haematopoietic culture conditions.

For both concentrations of Ren there is a dilution dependant increase of the Gr1+CD11b- population which may be a granulocyte subset (Fig.65b). Finally, in the ET-2 culture there is a reduction in B220-CD3+ cells which may include T cells and a marginal increase in the proportion of dendritic cells within the total live population.

The phenotype of haematopoietic cells was also analysed after 7 day culture of E10.5 AGMs for one experiment (Fig 65c). Haematopoietic output was much higher for this later stage culture as would be expected, with haematopoietic cells making up a higher proportion of the live cells than at E9.5. The increase in B cell production seen for E9.5 was not replicated at E10.5 except for ET-1 10. Production of Gr1+CD11b- cells were no longer seen in the Ren cultures but a significant population in ET-1 100. There was no evidence of the population termed lymphoid progenitor in the E10.5 cultures, suggesting that production of this population derives from earlier cells produced at E9.5.

Taken together, although addition of the each of the factors to culture does not have a statistically significant effect on any of the cell types analysed the data suggests that each factor drives a push towards an increase in haematopoietic output and that each factor may favour differentiation of certain lineages which may also be stage dependant.

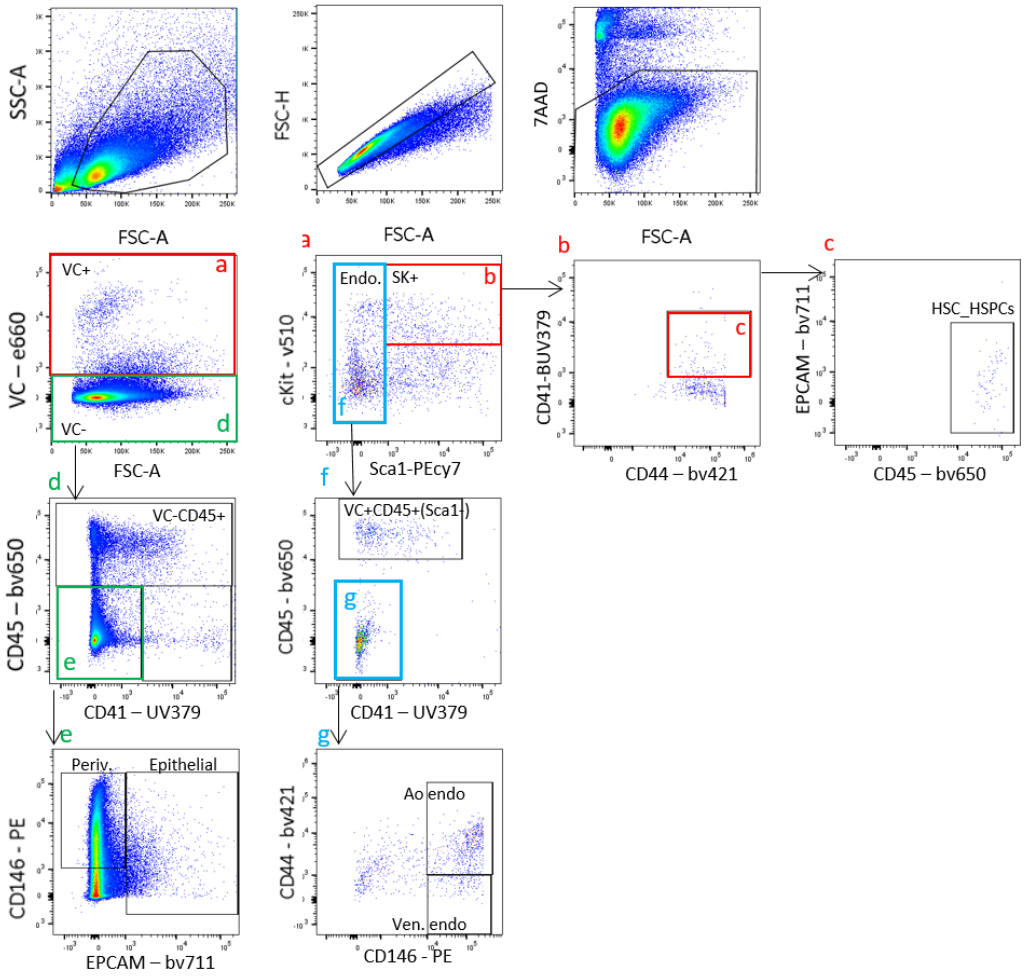


Figure 62. HSC_HSPC and vessel wall cell phenotype analysis of cultures after 7 days floating membrane culture. Gating strategy for Panel A. The top 3 plots shows gating for cells of interest, single cells and live cells respectively. Letters indicate parent populations and subsequent plots.

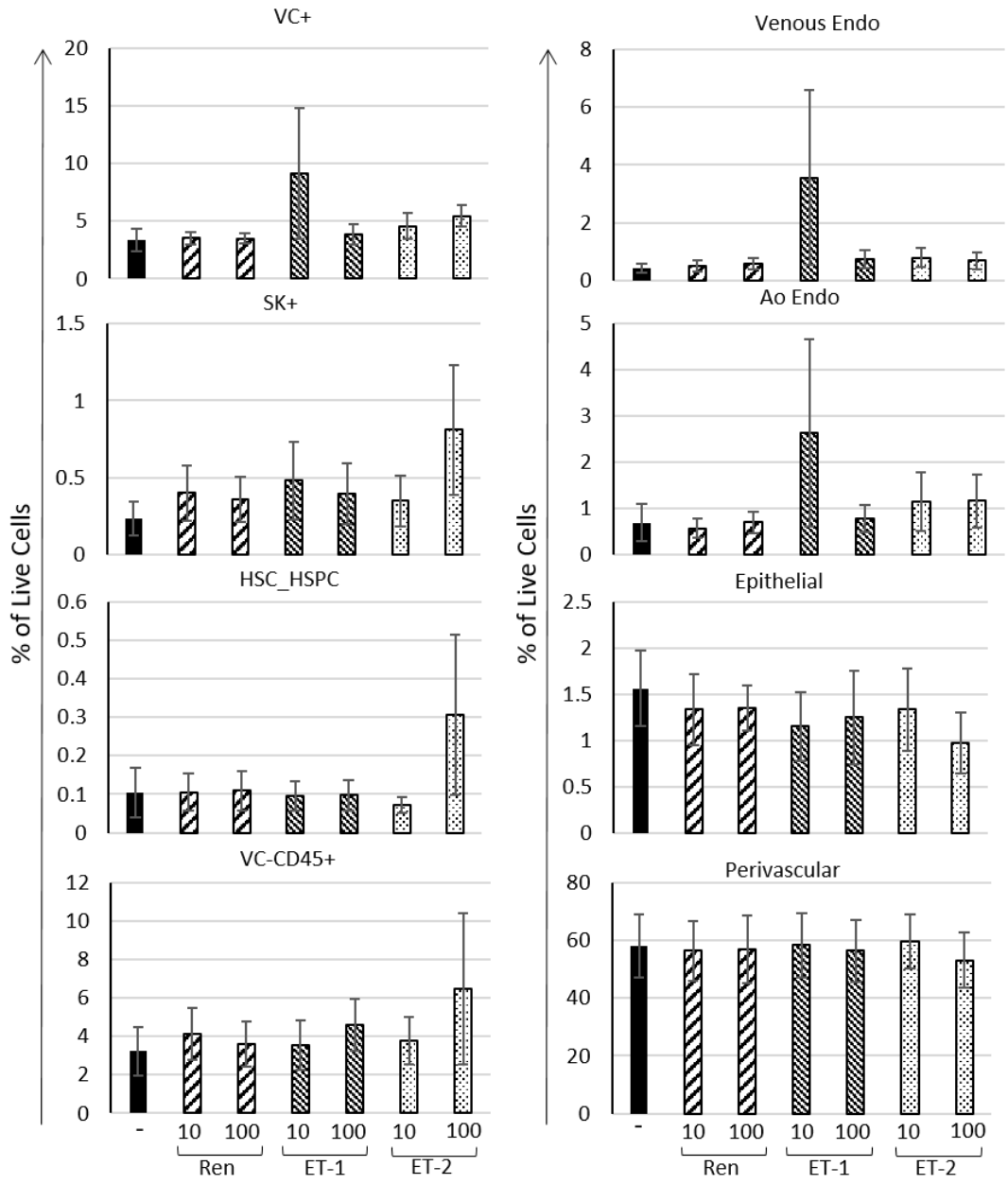


Figure 63. E9.5 phenotypic analysis of HSC_HSPC and vessel wall populations after 7 days floating membrane culture with Ren, ET-1 or ET-2 (10ng/µl and 100ng/µl each). N=4. Error bars are standard deviation around the mean. Paired t-test, $p<0.05$.

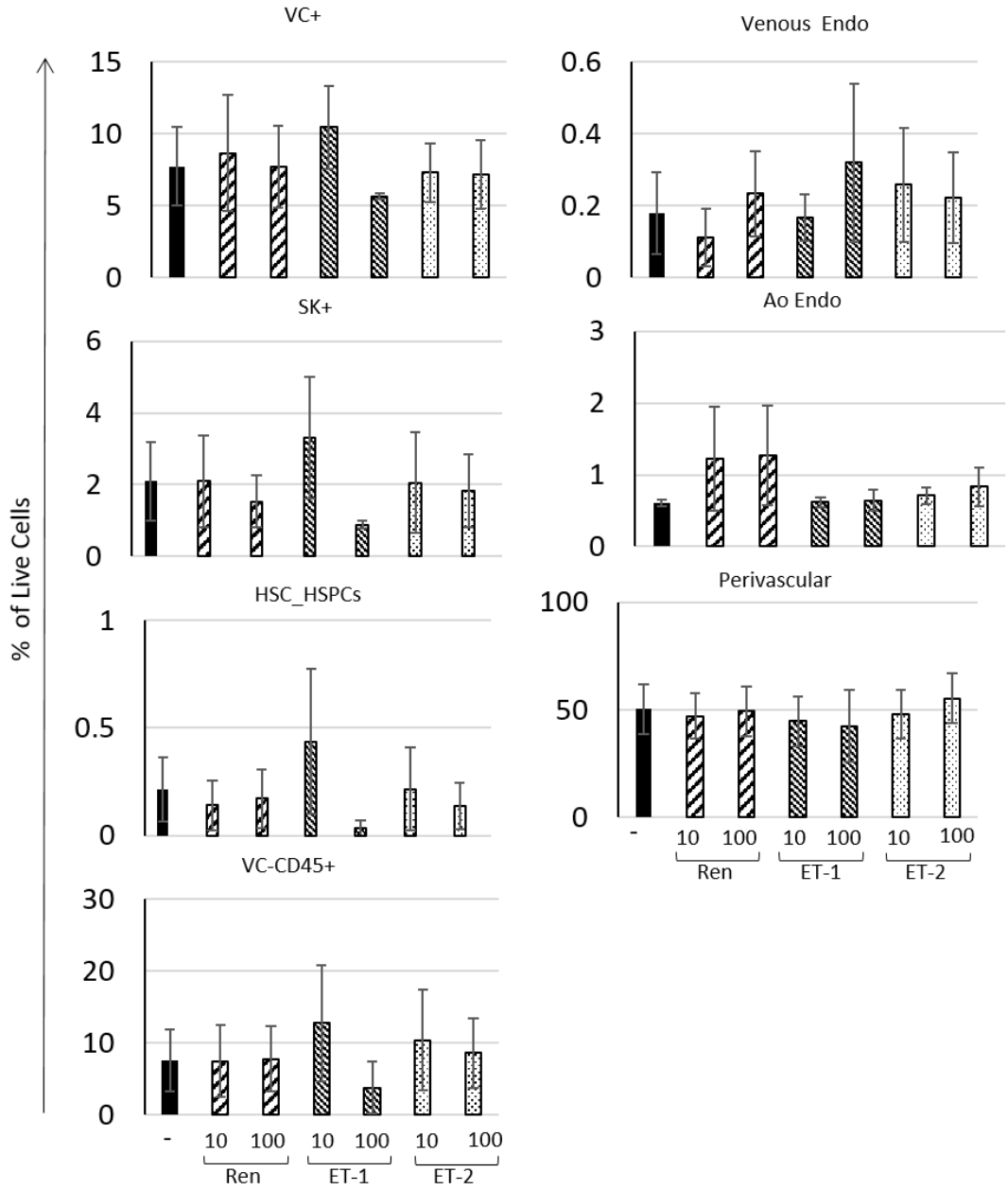


Figure 64. E10.5 phenotypic analysis of HSC_HSPC and vessel wall populations after 7 days floating membrane culture with Ren, ET-1 or ET-2 (10ng/µl and 100ng/µl each). N=3. Error bars are standard deviation around the mean.

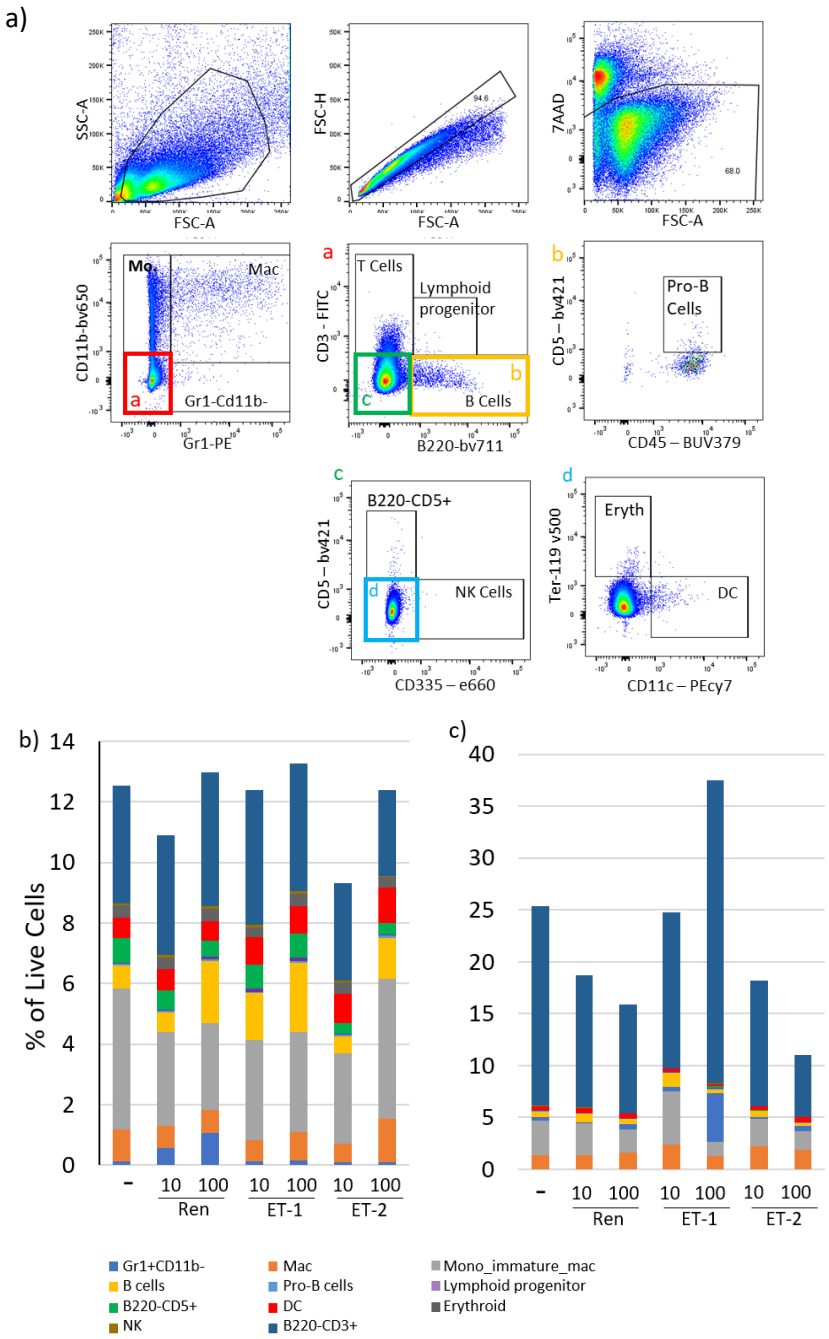


Figure 65: Haematopoietic lineage analysis of AGM cells after 7 day culture. a) Panel B gating strategy for analysis of E9.5 (n=3) (b) and E10.5 (n=1) (c) cells after 7 days culture as reaggregates and explants respectively. Proportions of haematopoietic cell types are shown as a percentage of live cells. Differences between test conditions are not significant (paired t.test, $p<0.05$).

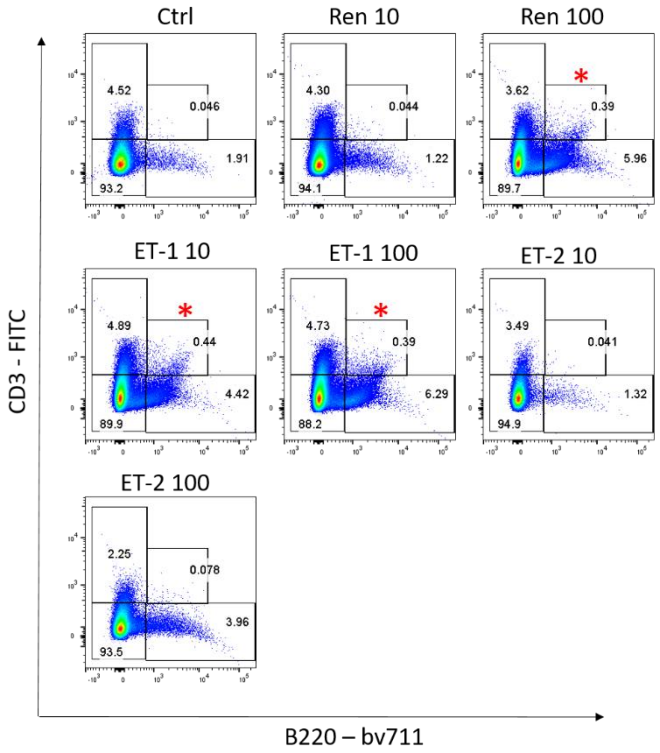


Figure 66. Example plots demonstrating appearance of a B220+CD3+ population for Ren 100, ET-1 10 and ET-100 conditions – indicated with an asterisk. The cell populations shown are from the Gr1-CD11b- population of live cells as depicted in Fig 64a.

5.2.3 *In vivo* transplantation assays demonstrate that Ren, ET-1 and ET-2 positively influence HSC production in E9.5 haematopoietic cultures.

Following 7 days E9.5 caudal tissue culture with either Ren, ET-1 or ET-2 cells were dissociated and transplanted into irradiated Ly5.1 mice. Donor repopulation (CD45_2) was assessed at 8 weeks and 18 weeks to confirm that repopulation was long-term. In each of 3 experiments the control gave no repopulation, which is likely due to the lack of support from OP9 cells in culture (Fig.67b,c)(Rybtsov *et al.*, 2011). In contrast each of the 3 test conditions gave repopulation (>5%) at both high and low concentrations which was significant ($p<0.05$, paired t-test) for both Ren 100 and ET-2 100. ET-2 100 had the highest number of long-term repopulated mice with 6/12 although Ren 100 had a similar number 5/12 but with higher reconstitution levels overall. ET-1 10 had more mice repopulated than ET-1 100 (3/12 compared to 1/12) which corresponds with ET-1 10 having a higher average SK+ population compared with ET-1 100 for both E9.5 and E10.5 cultures.

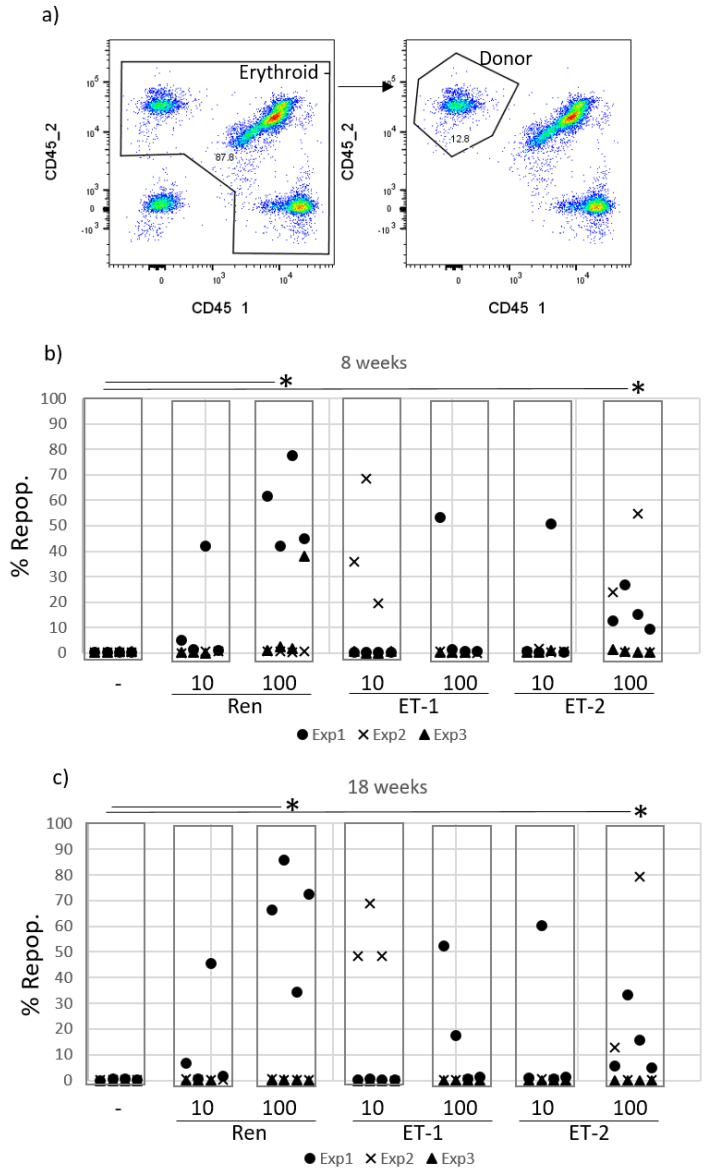


Figure 67 Addition of renin and endothelins increases the number of HSCs derived from E9.5 caudal part reaggregate cultures a) Example gating strategy for donor population CD45_2 cells from the peripheral blood of transplanted mice. Erythroid cells are first excluded (left panel). Single cells and live cells were gated as before. Short term (8 weeks) and Long-term (18 weeks) repopulation levels for each condition is shown in b) and c) respectively. An asterisk indicates statistically significant differences in the test condition compared with control (paired t-test, $p<0.05$).

At 20 weeks, mice were culled and multilineage analysis was carried out on haematopoietic organs to assess whether the repopulating HSCs gave multilineage reconstitution according to the gating strategy in Fig.68. Ren cultured HSCs reconstituted all lineages with a balance as previously described for E9.5 AGM cultures (Rybtsov *et al.*, 2014) with a dominant myeloid population in the bone marrow (BM), more balanced myeloid-lymphoid contribution in the blood and a lymphoid bias in the spleen (Fig.69a). ET-1 10 shows a similar distribution of lineages in the BM and spleen. In the blood however,

there is a greater proportion of T cells and minimal myeloid output. Conversely, ET-1 100 produces a drive for myeloid cell production in the BM, blood and spleen with almost no T cell or B cell production at all in the BM and very little in the blood. ET-2 has the opposite effect with a high lymphoid output in the BM and marginal myeloid cell production in blood. CD4 and CD8 T cell populations in the thymus were also analysed including CD4+CD8+ thymocytes and mature CD4+ helper and CD8+ cytotoxic T cells. Levels were equivalent for each condition (Fig.69b), indicating unimpaired homing of thymocytes to the thymus from the bone marrow and maturation.

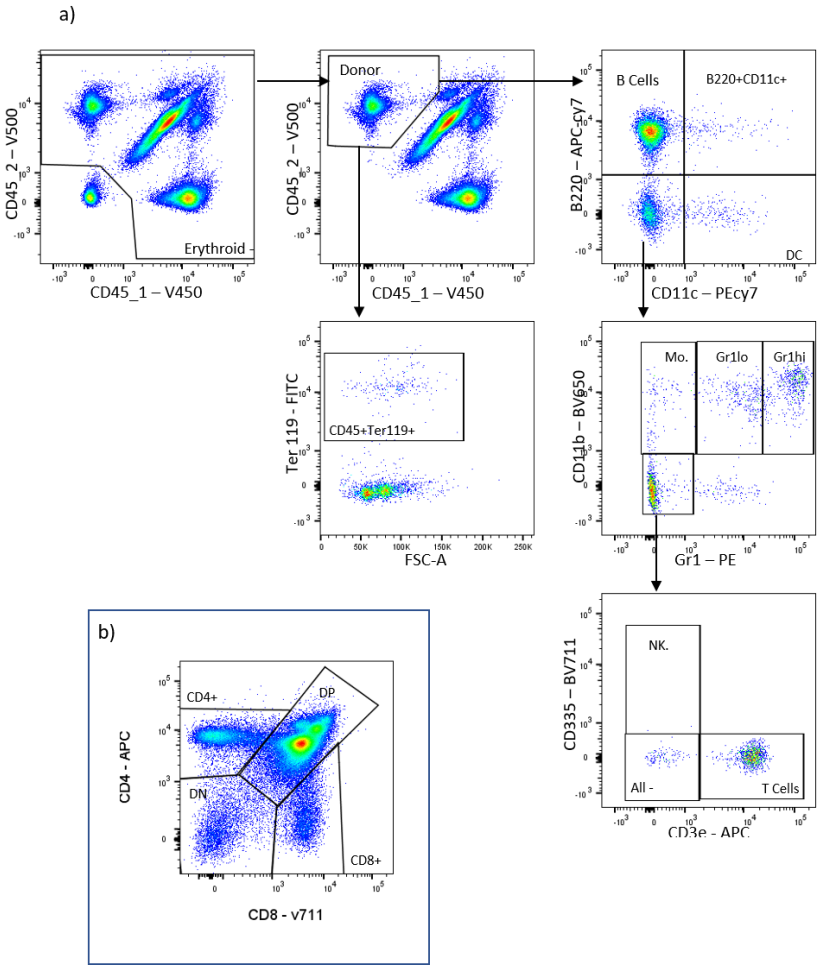


Figure 68. Gating strategy for analysis of multilineage reconstitution of mice transplanted with E9.5 reaggregate cultured CD45.1 cells. Bone marrow, blood spleen were analysed with panel (a) and thymus with panel (b). Plot in (b) is from donor cells gated in the same way as (a) and indicates double negative (DN), CD8+CD4+ double positive (DP) and CD4+, CD8+ thymocytes.

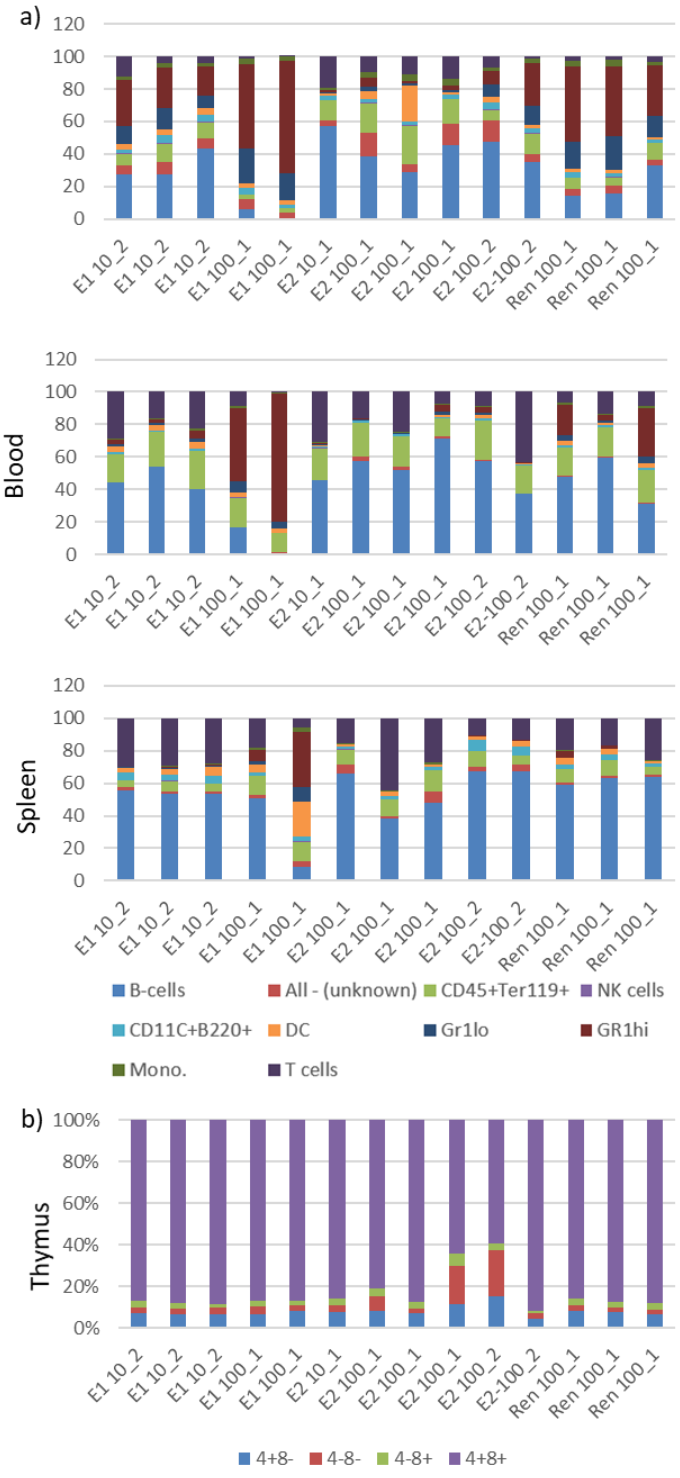


Figure 69. Multilineage analysis of transplanted mice at 20 weeks for two experiments: a) Relative proportion of haematopoietic cell subtypes in bone marrow (BM), blood and spleen. (b) Relative proportions of CD4/CD8 T cell sub-types in the thymus. Experiment number is indicated after the underscore for each sample along the x-axis.

5.3 **Discussion**

All 3 of the secreted factors investigated, renin, ET-1 and ET-2 have a positive effect on HSC generation from cells of the E9.5 embryonic caudal tissue as demonstrated by the repopulation assay. At E9.5 there are no pre-HSCs or adult-type HSCs yet present in the AGM, only pro-HSCs (Rybtsov *et al.*, 2014) which are matured to HSCs during the 7 day culture with cytokines. Culturing the reaggregates with OP9s provides the cells with a stabilising stroma and maturation is robust. However, OP9s were not used here in case they provided signals which confound the effects of the factors being tested. Indeed, datasets from (McGarvey *et al.*, 2017) show that Ren and ET-2 are expressed in OP9s at low levels and ET-1 at high levels. Without supporting OP9s the conditions were not supportive enough to derive HSCs from the control AGMs across 3 experiments. The reconstitution of mice from cultures with each renin, ET-1 and ET-2 indicate that these factors produced more favourable conditions for the production of HSCs with the strongest effects seen for renin and ET-2. It is unknown whether this is an enhanced maturation of pro-HSCs to adult-type HSCs, an increase in production of pro-HSCs from the haematogenic endothelial cells, or an expansion of pre-existing HSCs. More experiments will be required to investigate this. It may also be of use to co-culture the reaggregates with OP9s in order to get repopulation from the control cells and check that enhanced reconstitution in the test conditions holds true.

Intriguingly, ET-1 had contrasting effects at the lower and higher concentrations. At E9.5 ET-1 10 showed a slightly higher average SK+ population than ET-1 100 in the cultured cells which was even more exaggerated at E10.5 as well as in the HSC_HSPC and VC-CD45+ haematopoietic compartments. ET-1 100 in fact, had reduced SK+, HSC_HSPC and VC-CD45+ compartments compared with all other conditions. This suggests that the lower concentration may have a stimulatory effect on haematopoiesis which becomes inhibitory at the higher concentration. This is further evidenced by the higher levels of reconstitution in the repopulation assay by ET-1 10 over ET-1 100. Furthermore, the two concentrations appear to give different balance of multilineage reconstitution with ET-1 100 having a high myeloid bias, particularly in the bone marrow and blood compared to a more balanced reconstitution for ET-1 10.

The mouse model was used for these experiments due to the lack of availability of sufficient human samples for functional assays. Analysis of the previously published mouse AGM spatial transcriptome (McGarvey *et al.*, 2017) demonstrated conserved ventral enrichment of renin and endothelin pathway components indicating that a functional role for these factors may be conserved across mouse and human also. Furthermore, the close association of ET-1 protein expression with the IAHCs in the E10.5 mouse as was demonstrated in the human dorsal aorta in chapter 2 is further evidence of a conserved role between species. It is therefore likely that the stimulatory effect on the generation of HSCs

established in these experiments would be demonstrable with human cells. To this end human PSCs would be a good system for testing whether renin or endothelins can enhance generation of HSCs.

6. Concluding Remarks

The goal of this project was to generate an in-depth analysis of signalling within the human HSC developmental niche. Due to the relative lack of availability of early human embryos and the sensitivity of the material, signalling in the human AGM has thus far not been well characterised. Experiments in the mouse and other animal models have been highly informative towards our understanding of the processes and pathways underlying adult-type HSC emergence in the embryo. Although many mechanisms are conserved between model species and human, each organism has unique differences. This makes understanding the human AGM of great clinical importance as human-specific niche signals may be what are required to derive fully functional HSCs in culture.

Signalling across the AGM is polarised across the dorsal-ventral axis with both ventral and dorsal signals contributing towards promoting HSC emergence (Peeters *et al.*, 2009; Souilhol *et al.*, 2016a; McGarvey *et al.*, 2017). HSC emergence itself is polarised to the ventral domain of the dorsal aorta (Taoudi and Medvinsky, 2007b; Ivanovs *et al.*, 2011). This project set out to define the signals that cause HSCs to emerge in this spatio-temporally controlled manner. In order to interrogate the transcriptome of the dorsal aorta and surrounding cells without losing important spatial information, laser capture microdissection was used. The resulting spatial transcriptome provided a detailed map of signalling across the dorsal aorta region. Of particular use are the data revealing polarised genes and pathways across the dorsal-ventral axis. These polarised factors are the most likely candidates to have a role in regulating emergence of HSCs. In this study the genes *REN* and *EDN1* were chosen as the focus of functional assessment but this dataset can serve as the starting point for investigation of many other factors.

Renin, endothelin-1 and endothelin-2 were shown to positively influence HSC emergence through functional validation experiments which emphasises the utility of the datasets generated in this project. These experiments were in mouse and it is now important to begin functionally assessing these factors in a human model, namely human PSC cultures. Using the most current protocols to derive HSC-like cells from PSCs, these factors can be added to the culture at several time points i.e. mesoderm induction or haematogenic endothelium specification, to assess whether these factors can functionally improve the HSCs generated. Lessons from developmental biology have greatly facilitated attempts to derive HSCs artificially. These datasets are an additional contribution to the field to aid future investigations into the mechanisms behind HSC development and refine culture conditions.

7. References

Aamar, E. and Dawid, I. B. (2008) 'Protocadherin-18a has a role in cell adhesion, behavior and migration in zebrafish development', *Developmental Biology,* 318(2), pp. 335-346.

Ackermann, M., Ritthaler, T., Riegger, G., Kurtz, A. and Kramer, B. K. (1995) 'Endothelin inhibits cAMP-induced renin release from isolated renal juxtaglomerular cells', *J Cardiovasc Pharmacol,* 26 Suppl 3, pp. S135-7.

Adamo, L., Naveiras, O., Wenzel, P. L., McKinney-Freeman, S., Mack, P. J., Gracia-Sancho, J., Suchy-Dicey, A., Yoshimoto, M., Lensch, M. W., Yoder, M. C., García-Cardeña, G. and Daley, G. Q. (2009) 'Biomechanical forces promote embryonic haematopoiesis', *Nature,* 459(7250), pp. 1131.

Adolfsson, J., Mansson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C. T., Bryder, D., Yang, L., Borge, O. J., Thoren, L. A., Anderson, K., Sitnicka, E., Sasaki, Y., Sigvardsson, M. and Jacobsen, S. E. (2005) 'Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment', *Cell,* 121(2), pp. 295-306.

Akashi, K., Traver, D., Miyamoto, T. and Weissman, I. L. (2000) 'A clonogenic common myeloid progenitor that gives rise to all myeloid lineages', *Nature,* 404(6774), pp. 193-7.

Alexander, C., Zuniga, E., Blitz, I. L., Wada, N., Le Pabic, P., Javidan, Y., Zhang, T., Cho, K. W., Crump, J. G. and Schilling, T. F. (2011) 'Combinatorial roles for BMPs and Endothelin 1 in patterning the dorsal-ventral axis of the craniofacial skeleton', *Development,* 138(23), pp. 5135-46.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. . Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Armstrong, L., Lako, M., Buckley, N., Lappin, T. R., Murphy, M. J., Nolta, J. A., Pittenger, M. and Stojkovic, M. (2012) 'Editorial: Our top 10 developments in stem cell biology over the last 30 years', *Stem Cells,* 30(1), pp. 2-9.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G. (2000) 'Gene ontology: tool for the unification of biology. The Gene Ontology Consortium', *Nat Genet,* 25(1), pp. 25-9.

Ayllon, V., Bueno, C., Ramos-Mejia, V., Navarro-Montero, O., Prieto, C., Real, P. J., Romero, T., Garcia-Leon, M. J., Toribio, M. L., Bigas, A. and Menendez, P. (2015) 'The Notch ligand DLL4 specifically marks human hematoendothelial progenitors and regulates their hematopoietic fate', *Leukemia,* 29(8), pp. 1741-53.

Baron, C. S., Kester, L., Klaus, A., Boisset, J. C., Thambyrajah, R., Yvernogeau, L., Kouskoff, V., Lacaud, G., van Oudenaarden, A. and Robin, C. (2018) 'Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta', *Nat Commun,* 9(1), pp. 2517.

Batard, P., Monier, M. N., Fortunel, N., Ducos, K., Sansilvestri-Morel, P., Phan, T. H., Hatzfeld, A. and Hatzfeld, J. A. (2000) 'TGF-beta 1 maintains hematopoietic immaturity by a reversible negative control of cell cycle and induces CD34 antigen up-modulation', *Journal of Cell Science,* 113(3), pp. 383-390.

Batsivari, A., Rybtsov, S., Souilhol, C., Binagui-Casas, A., Hills, D., Zhao, S., Travers, P. and Medvinsky, A. (2017) 'Understanding Hematopoietic Stem Cell Development through Functional Correlation of Their Proliferative Status with the Intra-aortic Cluster Architecture', *Stem Cell Reports,* 8(6), pp. 1549-1562.

Becker, A., McCulloch, E. A. and Till, J. E. 1963. Cytological Demonstration of the Clonal Nature of Spleen Colonies Derived from Transplanted Mouse Marrow Cells. Nature.

Beierwaltes, W. H. and Carretero, O. A. (1992) 'Nonprostanoid endothelium-derived factors inhibit renin release', *Hypertension,* 19(2 Suppl), pp. Ii68-73.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society. Series B (Methodological),* 57(1), pp. 289-300.

Berglund, L. M., Lyssenko, V., Ladenvall, C., Kotova, O., Edsfeldt, A., Pilgaard, K., Alkayyali, S., Brons, C., Forsblom, C., Jonsson, A., Zetterqvist, A. V., Nitulescu, M., McDavitt, C. R., Duner, P., Stancakova, A., Kuusisto, J., Ahlqvist, E., Lajer, M., Tarnow, L., Madsbad, S., Rossing, P., Kieffer, T. J., Melander, O., Orho-Melander, M., Nilsson, P., Groop, P. H., Vaag, A., Lindblad, B., Gottsater, A., Laakso, M., Goncalves, I., Groop, L. and Gomez, M. F. (2016) 'Glucose-Dependent Insulinotropic Polypeptide Stimulates Osteopontin Expression in the Vasculature via Endothelin-1 and CREB', *Diabetes,* 65(1), pp. 239-54.

Bertrand, J. Y., Chi, N. C., Santoso, B., Teng, S. T., Stainier, D. Y. R. and Traver, D. (2010) 'Haematopoietic stem cells derive directly from aortic endothelium during development', *Nature,* 464(7285), pp. 108-U120.

Bertrand, J. Y., Jalil, A., Klaine, M., Jung, S., Cumano, A. and Godin, I. (2005) 'Three pathways to mature macrophages in the early mouse yolk sac', *Blood,* 106(9), pp. 3004-11.

Blighe, K. 2018. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. https://github.com/kevinblighe .

Bloom, W. and Bartelmez, G. W. (1940) 'Hematopoiesis in young human embryos', *American Journal of Anatomy,* 67(1), pp. 21-53.

Boiers, C., Carrelha, J., Lutteropp, M., Luc, S., Green, J. C., Azzoni, E., Woll, P. S., Mead, A. J., Hultquist, A., Swiers, G., Perdiguero, E. G., Macaulay, I. C., Melchiori, L., Luis, T. C., Kharazi, S., Bouriez-Jones, T., Deng, Q., Ponten, A., Atkinson, D., Jensen, C. T., Sitnicka, E., Geissmann, F., Godin, I., Sandberg, R., de Bruijn, M. F. and Jacobsen, S. E. (2013) 'Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells', *Cell Stem Cell,* 13(5), pp. 535-48.

Boisset, J. C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E. and Robin, C. (2010) 'In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium', *Nature,* 464(7285), pp. 116-U131.

Brennecke, P., Anders, S., Kim, J. K., Kołodziejczyk, A. A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S. A., Marioni, J. C. and Heisler, M. G. (2013) 'Accounting for technical noise in single-cell RNA-seq experiments', *Nature Methods,* 10(11), pp. 1093.

BronnerFraser, M. and Fraser, S. E. (1997) 'Differentiation of the vertebrate neural tube', *Current Opinion in Cell Biology,* 9(6), pp. 885-891.

Buchbinder, E. I. and Desai, A. (2016) 'CTLA-4 and PD-1 Pathways: Similarities, Differences, and Implications of Their Inhibition', *Am J Clin Oncol,* 39(1), pp. 98-106.

Burns, C. E., Traver, D., Mayhall, E., Shepard, J. L. and Zon, L. I. (2005) 'Hematopoietic stem cell fate is established by the Notch-Runx pathway', *Genes Dev,* 19(19), pp. 2331-42.

Cabezas-Wallscheid, N., Klimmeck, D., Hansson, J., Lipka, D. B., Reyes, A., Wang, Q., Weichenhan, D., Lier, A., von Paleske, L., Renders, S., Wunsche, P., Zeisberger, P., Brocks, D., Gu, L., Herrmann, C., Haas, S., Essers, M. A. G., Brors, B., Eils, R., Huber, W., Milsom, M. D., Plass, C., Krijgsveld, J. and Trumpp, A. (2014) 'Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis', *Cell Stem Cell,* 15(4), pp. 507-522.

Campbell, K. and Casanova, J. (2016) 'A common framework for EMT and collective cell migration'.

Capurro, M., Martin, T., Shi, W. and Filmus, J. (2014) 'Glypican-3 binds to Frizzled and plays a direct role in the stimulation of canonical Wnt signaling', *J Cell Sci,* 127(Pt 7), pp. 1565-75.

Capurro, M. I., Xu, P., Shi, W., Li, F., Jia, A. and Filmus, J. (2008) 'Glypican-3 inhibits Hedgehog signaling during development by competing with patched for Hedgehog binding', *Dev Cell,* 14(5), pp. 700-11.

Chadwick, K., Wang, L., Li, L., Menendez, P., Murdoch, B., Rouleau, A. and Bhatia, M. (2003) 'Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells', *Blood,* 102(3), pp. 906-15.

Chanda, B., Ditadi, A., Iscove, N. N. and Keller, G. (2013) 'Retinoic acid signaling is essential for embryonic hematopoietic stem cell development', *Cell,* 155(1), pp. 215-27.

Chen, M. J., Li, Y., De Obaldia, M. E., Yang, Q., Yzaguirre, A. D., Yamada-Inagawa, T., Vink, C. S., Bhandoola, A., Dzierzak, E. and Speck, N. A. (2011) 'Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells', *Cell Stem Cell,* 9(6), pp. 541-52.

Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. and Speck, N. A. (2009) 'Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter', *Nature,* 457(7231), pp. 887-891.

Chua, B. H. L., Chua, C. C., Diglio, C. A. and Siu, B. B. (1993) 'REGULATION OF ENDOTHELIN-1 MESSENGER-RNA BY ANGIOTENSIN-II IN RAT-HEART ENDOTHELIAL-CELLS', *Biochimica Et Biophysica Acta,* 1178(2), pp. 201-206.

Ciau-Uitz, A., Pinheiro, P., Gupta, R., Enver, T. and Patient, R. (2010) 'Tel1/ETV6 specifies blood stem cells through the agency of VEGF signaling', *Dev Cell,* 18(4), pp. 569-78.

Ciau-Uitz, A., Pinheiro, P., Kirmizitas, A., Zuo, J. and Patient, R. (2013) 'VEGFA-dependent and -independent pathways synergise to drive Scl expression and initiate programming of the blood stem cell lineage in Xenopus', *Development,* 140(12), pp. 2632-42.

Ciau-Uitz A, P. R., Medvinsky A (2016) 'Ontogeny of the Hematopoietic System', *Encyclopedia of Immunobiology*: Elsevier, pp. 1 - 14.

Ciau-Uitz, A., Walmsley, M. and Patient, R. (2000) 'Distinct origins of adult and embryonic blood in Xenopus', *Cell,* 102(6), pp. 787-96.

Clarke, R. L., Yzaguirre, A. D., Yashiro-Ohtani, Y., Bondue, A., Blanpain, C., Pear, W. S., Speck, N. A. and Keller, G. (2013) 'The expression of Sox17 identifies and regulates haemogenic endothelium', *Nature Cell Biology,* 15(5), pp. 502-+.

Clements, W. K., Kim, A. D., Ong, K. G., Moore, J. C., Lawson, N. D. and Traver, D. (2011) 'A somitic Wnt16/Notch pathway specifies haematopoietic stem cells', *Nature,* 474(7350), pp. 220-4.

Cohen, S., Roy, J., Lachance, S., Delisle, J. S., Marinier, A., Busque, L., Roy, D. C., Barabe, F., Ahmad, I., Bambace, N., Bernard, L., Kiss, T., Bouchard, P., Caudrelier, P., Landais, S., Larochelle, F., Chagraoui, J., Lehnertz, B., Corneau, S., Tomellini, E., van Kampen, J. J. A., Cornelissen, J. J., Dumont-Lagace, M.,

Tanguay, M., Li, Q., Lemieux, S., Zandstra, P. W. and Sauvageau, G. (2020) 'Hematopoietic stem cell transplantation using single UM171-expanded cord blood: a single-arm, phase 1-2 safety and feasibility study', *Lancet Haematol,* 7(2), pp. e134-e145.

Cole, J., Ertoy, D., Lin, H. C., Sutliff, R. L., Ezan, E., Guyene, T. T., Capecchi, M., Corvol, P. and Bernstein, K. E. (2000) 'Lack of angiotensin II-facilitated erythropoiesis causes anemia in angiotensin-converting enzyme-deficient mice', *Journal of Clinical Investigation,* 106(11), pp. 1391-1398.

Corada, M., Orsenigo, F., Morini, M. F., Pitulescu, M. E., Bhat, G., Nyqvist, D., Breviario, F., Conti, V., Briot, A., Iruela-Arispe, M. L., Adams, R. H. and Dejana, E. (2013) 'Sox17 is indispensable for acquisition and maintenance of arterial identity', *Nat Commun,* 4, pp. 2609.

Crisan, M., Solaimani Kartalaei, P., Neagu, A., Karkanpouna, S., Yamada-Inagawa, T., Purini, C., Vink, C., van der Linden, R., van Ijcken, W., Chuva de Sousa Lopes, S., Monteiro, R., Mummery, C. and Dzierzak, E. (2016) 'BMP and Hedgehog Regulate Distinct AGM Hematopoietic Stem Cells Ex Vivo', *Stem Cell Reports: Vol. 3*, pp. 383-95.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P. and Stein, L. (2011) 'Reactome: a database of reactions, pathways and biological processes', *Nucleic Acids Res: Vol. Database issue*, pp. D691-7.

D'Souza A, F. C. (2018) 'Current Uses and Outcomes of Hematopoietic Cell Transplantation (HCT): CIBMTR Summary Slides'.

Damm, E. W. and Clements, W. K. (2017) 'Pdgf signalling guides neural crest contribution to the haematopoietic stem cell specification niche', *Nature Cell Biology,* 19(5), pp. 457-+.

Daniel, M. G., Pereira, C. F., Lemischka, I. R. and Moore, K. A. (2016) 'Making a Hematopoietic Stem Cell', *Trends in Cell Biology,* 26(3), pp. 202-214.

de Bruijn, M. F., Speck, N. A., Peeters, M. C. and Dzierzak, E. (2000) 'Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo', *Embo j,* 19(11), pp. 2465-74.

de Pater, E., Kaimakis, P., Vink, C. S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P. S., Camper, S. A., Speck, N. and Dzierzak, E. (2013) 'Gata2 is required for HSC generation and survival', *J Exp Med,* 210(13), pp. 2843-50.

Deuel, T. F., Zhang, N., Yeh, H. J., Silos-Santiago, I. and Wang, Z. Y. (2002) 'Pleiotrophin: a cytokine with diverse functions and a novel signaling pathway', *Arch Biochem Biophys,* 397(2), pp. 162-71.

Diaz, M. F., Li, N., Lee, H. J., Adamo, L., Evans, S. M., Willey, H. E., Arora, N., Torisawa, Y., Vickers, D. A., Morris, S. A., Naveiras, O., Murthy, S. K., Ingber, D. E., Daley, G. Q., Garcia-Cardena, G. and Wenzel, P. L. (2015) 'Biomechanical forces promote blood development through prostaglandin E-2 and the cAMP-PKA signaling axis', *Journal of Experimental Medicine,* 212(5), pp. 665-680.

Dieterlen-Lievre, F. (1975) 'On the origin of haemopoietic stem cells in the avian embryo: an experimental approach', *J Embryol Exp Morphol,* 33(3), pp. 607-19.

Dinarello, C. A. (2000) 'Proinflammatory cytokines', *Chest,* 118(2), pp. 503-8.

Discher, D. E., Mooney, D. J. and Zandstra, P. W. (2009) 'Growth factors, matrices, and forces combine and control stem cells', *Science,* 324(5935), pp. 1673-7.

Ditadi, A., Sturgeon, C. M. and Keller, G. (2016) 'A view of human haematopoietic development from the Petri dish', *Nature Reviews Molecular Cell Biology,* 18(1), pp. 56-67.

Ditadi, A., Sturgeon, C. M., Tober, J., Awong, G., Kennedy, M., Yzaguirre, A. D., Azzola, L., Ng, E. S., Stanley, E. G., French, D. L., Cheng, X., Gadue, P., Speck, N. A., Elefanty, A. G. and Keller, G. (2015) 'Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages', *Nat Cell Biol,* 17(5), pp. 580-91.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics,* 29(1), pp. 15-21.

Dodt, M., Roehr, J. T., Ahmed, R. and Dieterich, C. (2012) 'FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms', *Biology (Basel),* 1(3), pp. 895-905.

Dooley, K. A., Davidson, A. J. and Zon, L. I. (2005) 'Zebrafish scl functions independently in hematopoietic and endothelial development', *Dev Biol,* 277(2), pp. 522-36.

Dou, D. R., Calvanese, V., Sierra, M. I., Nguyen, A. T., Minasian, A., Saarikoski, P., Sasidharan, R., Ramirez, C. M., Zack, J. A., Crooks, G. M., Galic, Z. and Mikkola, H. K. (2016) 'Medial HOXA genes demarcate haematopoietic stem cell fate during human development', *Nat Cell Biol,* 18(6), pp. 595-606.

Doulatov, S., Notta, F., Eppert, K., Nguyen, L. T., Ohashi, P. S. and Dick, J. E. (2010) 'Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development', *Nature Immunology,* 11(7), pp. 585.

Doulatov, S., Vo, L. T., Chou, S. S., Kim, P. G., Arora, N., Li, H., Hadland, B. K., Bernstein, I. D., Collins, J. J., Zon, L. I. and Daley, G. Q. (2013) 'Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors', *Cell Stem Cell,* 13(4), pp. 459-70.

Drukker, M. and Benvenisty, N. (2004) 'The immunogenicity of human embryonic stem-derived cells', *Trends Biotechnol,* 22(3), pp. 136-41.

Duprez, D. A. (2006) 'Role of the renin-angiotensin-aldosterone system in vascular remodeling and inflammation: a clinical review', *J Hypertens,* 24(6), pp. 983-91.

Dzierzak, E. and Bigas, A. (2018) 'Blood Development: Hematopoietic Stem Cell Dependence and Independence', *Cell Stem Cell,* 22(5), pp. 639-651.

Easterbrook, J., Rybtsov, S., Gordon-Keylock, S., Ivanovs, A., Taoudi, S., Anderson, R. A. and Medvinsky, A. (2019) 'Analysis of the Spatiotemporal Development of Hematopoietic Stem and Progenitor Cells in the Early Human Embryo', *Stem Cell Reports,* 12(5), pp. 1056-1068.

Echelard, Y., Epstein, D. J., Stjacques, B., Shen, L., Mohler, J., McMahon, J. A. and McMahon, A. P. (1993) 'SONIC-HEDGEHOG, A MEMBER OF A FAMILY OF PUTATIVE SIGNALING MOLECULES, IS IMPLICATED IN THE REGULATION OF CNS POLARITY', *Cell,* 75(7), pp. 1417-1430.

Eich, C., Arlt, J., Vink, C. S., Solaimani Kartalaei, P., Kaimakis, P., Mariani, S. A., van der Linden, R., van Cappellen, W. A. and Dzierzak, E. (2018) 'In vivo single cell analysis reveals Gata2 dynamics in cells transitioning to hematopoietic fate', *J Exp Med,* 215(1), pp. 233-248.

Eilken, H. M., Nishikawa, S. and Schroeder, T. (2009) 'Continuous single-cell imaging of blood generation from haemogenic endothelium', *Nature,* 457(7231), pp. 896-900.

Ema, H. and Nakauchi, H. (2000) 'Expansion of hematopoietic stem cells in the developing liver of a mouse embryo', *Blood,* 95(7), pp. 2284-8.

Espin-Palazon, R., Stachura, D. L., Campbell, C. A., Garcia-Moreno, D., Del Cid, N., Kim, A. D., Candel, S., Meseguer, J., Mulero, V. and Traver, D. (2014) 'Proinflammatory Signaling Regulates Hematopoietic Stem Cell Emergence', *Cell,* 159(5), pp. 1070-1085.

European Society for Bone Marrow Transplantation - Annual Report 2018 (2018).

Evers, D. L., Fowler, C. B., Cunningham, B. R., Mason, J. T. and O'Leary, T. J. (2011) 'The Effect of Formaldehyde Fixation on RNA Optimization of Formaldehyde Adduct Removal', *Journal of Molecular Diagnostics,* 13(3), pp. 282-288.

Feng, Q., Chai, C., Jiang, X. S., Leong, K. W. and Mao, H. Q. (2006) 'Expansion of engrafting human hematopoietic stem/progenitor cells in three-dimensional scaffolds with surface-immobilized fibronectin', *J Biomed Mater Res A,* 78(4), pp. 781-91.

Ferkowicz, M. J. and Yoder, M. C. (2005) 'Blood island formation: longstanding observations and modern interpretations', *Exp Hematol,* 33(9), pp. 1041-7.

Fitch, S. R., Kimber, G. M., Wilson, N. K., Parker, A., Mirshekar-Syahkal, B., Gottgens, B., Medvinsky, A., Dzierzak, E. and Ottersbach, K. (2012) 'Signaling from the sympathetic nervous system regulates hematopoietic stem cell emergence during embryogenesis', *Cell Stem Cell,* 11(4), pp. 554-66.

Foudi, A., Hochedlinger, K., Van Buren, D., Schindler, J. W., Jaenisch, R., Carey, V. and Hock, H. (2009) 'Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells', *Nat Biotechnol,* 27(1), pp. 84-90.

Fukuda, T. (1973) 'Fetal hemopoiesis. I. Electron microscopic studies on human yolk sac hemopoiesis', *Virchows Arch B Cell Pathol,* 14(3), pp. 197-213.

Furumoto, T. A., Miura, N., Akasaka, T., Mizutani-Koseki, Y., Sudo, H., Fukuda, K., Maekawa, M., Yuasa, S., Fu, Y., Moriya, H., Taniguchi, M., Imai, K., Dahl, E., Balling, R., Pavlova, M., Gossler, A. and Koseki, H. (1999) 'Notochord-dependent expression of MFH1 and PAX1 cooperates to maintain the proliferation of sclerotome cells during the vertebral column development', *Dev Biol,* 210(1), pp. 15-29.

Garbe, A., Spyridonidis, A., Mobest, D., Schmoor, C., Mertelsmann, R. and Henschler, R. (1997) 'Transforming growth factor-beta 1 delays formation of granulocyte-macrophage colony-forming cells, but spares more primitive progenitors during ex vivo expansion of CD34(+) haemopoietic progenitor cells', *British Journal of Haematology,* 99(4), pp. 951-958.

Garcia-Porrero, J. A., Godin, I. E. and Dieterlen-Lievre, F. (1995) 'Potential intraembryonic hemogenic sites at pre-liver stages in the mouse', *Anat Embryol (Berl),* 192(5), pp. 425-35.

Gekas, C., Dieterlen-Lievre, F., Orkin, S. H. and Mikkola, H. K. (2005) 'The placenta is a niche for hematopoietic stem cells', *Dev Cell,* 8(3), pp. 365-75.

Gering, M. and Patient, R. (2005) 'Hedgehog signaling is required for adult blood stem cell formation in zebrafish embryos', *Dev Cell,* 8(3), pp. 389-400.

Gering, M., Rodaway, A. R., Göttgens, B., Patient, R. K. and Green, A. R. (1998) 'The SCL gene specifies haemangioblast development from early mesoderm', *EMBO J,* 17(14), pp. 4029-45.

Giampaolo, S., Wojcik, G., Serfling, E. and Patra, A. K. (2017) 'Interleukin-2-regulatory T cell axis critically regulates maintenance of hematopoietic stem cells', *Oncotarget,* 8(18), pp. 29625-29642.

Gilbert, S. F. (2000) 'The Neural Crest'.

Ginhoux, F. and Guilliams, M. (2016) 'Tissue-Resident Macrophage Ontogeny and Homeostasis', *Immunity,* 44(3), pp. 439-449.

Goessling, W., North, T. E., Loewer, S., Lord, A. M., Lee, S., Stoick-Cooper, C. L., Weidinger, G., Puder, M., Daley, G. Q., Moon, R. T. and Zon, L. I. (2009) 'Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration', *Cell,* 136(6), pp. 1136-47.

Gordon-Keylock, S., Sobiesiak, M., Rybtsov, S., Moore, K. and Medvinsky, A. (2013) 'Mouse extraembryonic arterial vessels harbor precursors capable of maturing into definitive HSCs', *Blood,* 122(14), pp. 2338-45.

Grainger, S., Richter, J., Palazon, R. E., Pouget, C., Lonquich, B., Wirth, S., Grassme, K. S., Herzog, W., Swift, M. R., Weinstein, B. M., Traver, D. and Willert, K. (2016) 'Wnt9a Is Required for the Aortic Amplification of Nascent Hematopoietic Stem Cells', *Cell Reports,* 17(6), pp. 1595-1606.

Guibentif, C. and Göttgens, B. (2017) 'Blood: Education for stem cells', *Nature,* 545(7655), pp. 415.

Guidi, N., Sacma, M., Standker, L., Soller, K., Marka, G., Eiwen, K., Weiss, J. M., Kirchhoff, F., Weil, T., Cancelas, J. A., Florian, M. C. and Geiger, H. (2017) 'Osteopontin attenuates aging-associated phenotypes of hematopoietic stem cells', *Embo j,* 36(7), pp. 840-853.

Hahn, A. W. A., Resink, T. J., Scottburden, T., Powell, J., Dohi, Y. and Buhler, F. R. (1990) 'STIMULATION OF ENDOTHELIN MESSENGER-RNA AND SECRETION IN RAT VASCULAR SMOOTH-MUSCLE CELLS - A NOVEL AUTOCRINE FUNCTION', *Cell Regulation,* 1(9), pp. 649-659.

Haque, S.-u., Welch, H. and Loizidou, M. (2013) *Handbook of Biologically Active Peptides.* Second edn.: Academic Press, p. 512-518.

Heissig, B., Lund, L. R., Akiyama, H., Ohki, M., Morita, Y., Rømer, J., Nakauchi, H., Okumura, K., Ogawa, H., Werb, Z., Danø, K. and Hattori, K. (2007) 'The Plasminogen Fibrinolytic Pathway Is Required for Hematopoietic Regeneration', *Cell Stem Cell,* 1(6), pp. 658-70.

Himburg, H. A., Termini, C. M., Schlussel, L., Kan, J., Li, M., Zhao, L., Fang, T., Sasine, J. P., Chang, V. Y. and Chute, J. P. (2018) 'Distinct Bone Marrow Sources of Pleiotrophin Control Hematopoietic Stem Cell Maintenance and Regeneration', *Cell Stem Cell,* 23(3), pp. 370-381.e5.

Houssaint, E. (1981) 'Differentiation of the mouse hepatic primordium. II. Extrinsic origin of the haemopoietic cell line', *Cell Differ,* 10(5), pp. 243-52.

Hu, K., Yu, J., Suknuntha, K., Tian, S., Montgomery, K., Choi, K. D., Stewart, R., Thomson, J. A. and Slukvin, II (2011) 'Efficient generation of transgene-free induced pluripotent stem cells from normal and neoplastic bone marrow and cord blood mononuclear cells', *Blood,* 117(14), pp. e109-19.

Huang, Y., Wongamorntham, S., Kasting, J., McQuillan, D., Owens, R. T., Yu, L., Noble, N. A. and Border, W. (2006) 'Renin increases mesangial cell transforming growth factor-beta 1 and matrix proteins through receptor-mediated, angiotensin II-independent mechanisms', *Kidney International,* 69(1), pp. 105-113.

Iizuka, K., Yokomizo, T., Watanabe, N., Tanaka, Y., Osato, M., Takaku, T. and Komatsu, N. (2016) 'Lack of Phenotypical and Morphological Evidences of Endothelial to Hematopoietic Transition in the Murine Embryonic Head during Hematopoietic Stem Cell Emergence', *PLoS One,* 11(5), pp. e0156427.

Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C. and Teichmann, S. A. (2016) 'Classification of low quality cells from single-cell RNA-seq data', *Genome Biol*.

Imai, T., Hirata, Y., Emori, T., Yanagisawa, M., Masaki, T. and Marumo, F. (1992) 'INDUCTION OF ENDOTHELIN-1 GENE BY ANGIOTENSIN AND VASOPRESSIN IN ENDOTHELIAL-CELLS', *Hypertension,* 19(6), pp. 753-757.

Ishitobi, H., Matsumoto, K., Azami, T., Itoh, F., Itoh, S., Takahashi, S. and Ema, M. (2010) 'Flk1-GFP BAC Tg mice: an animal model for the study of blood vessel development', *Exp Anim,* 59(5), pp. 615-22.

Ismailoglu, I., Yeamans, G., Daley, G. Q., Perlingeiro, R. C. and Kyba, M. (2008) 'Mesodermal patterning activity of SCL', *Exp Hematol,* 36(12), pp. 1593-603.

Ivanovs, A., Rybtsov, S., Anderson, R. A., Turner, M. L. and Medvinsky, A. (2014) 'Identification of the Niche and Phenotype of the First Human Hematopoietic Stem Cells', *Stem Cell Reports,* 2(4), pp. 449-456.

Ivanovs, A., Rybtsov, S., Ng, E. S., Stanley, E. G., Elefanty, A. G. and Medvinsky, A. (2017) 'Human haematopoietic stem cell development: from the embryo to the dish'.

Ivanovs, A., Rybtsov, S., Welch, L., Anderson, R. A., Turner, M. L. and Medvinsky, A. (2011) 'Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region', *Journal of Experimental Medicine,* 208(12), pp. 2417-2427.

Jacobson, L. O., Simmons, E. L., Marks, E. K. and Eldredge, J. H. (1951) 'Recovery from Radiation Injury'.

Jaffredo, T., Gautier, R., Eichmann, A. and Dieterlen-Lievre, F. (1998) 'Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny', *Development,* 125(22), pp. 4575-4583.

Johnson, G. R. and Moore, M. A. (1975) 'Role of stem cell migration in initiation of mouse foetal liver haemopoiesis', *Nature,* 258(5537), pp. 726-8.

Jokubaitis, V. J., Sinka, L., Driessen, R., Whitty, G., Haylock, D. N., Bertoncello, I., Smith, I., Peault, B., Tavian, M. and Simmons, P. J. (2008) 'Angiotensin-converting enzyme (CD143) marks hematopoietic stem cells in human embryonic, fetal, and adult hematopoietic tissues', *Blood,* 111(8), pp. 4055-4063.

Kanehisa, M. and Goto, S. (2000) 'KEGG: kyoto encyclopedia of genes and genomes', *Nucleic Acids Res,* 28(1), pp. 27-30.

Kang, H., Mesquitta, W. T., Jung, H. S., Moskvin, O. V., Thomson, J. A. and Slukvin, II (2018) 'GATA2 Is Dispensable for Specification of Hemogenic Endothelium but Promotes Endothelial-to-Hematopoietic Transition', *Stem Cell Reports,* 11(1), pp. 197-211.

Kennedy, M., Awong, G., Sturgeon, C. M., Ditadi, A., LaMotte-Mohs, R., Zuniga-Pflucker, J. C. and Keller, G. (2012) 'T lymphocyte potential marks the emergence of definitive hematopoietic progenitors in human pluripotent stem cell differentiation cultures', *Cell Rep,* 2(6), pp. 1722-35.

Kissa, K. and Herbomel, P. (2010) 'Blood stem cells emerge from aortic endothelium by a novel type of cell transition', *Nature,* 464(7285), pp. 112-U125.

Kissa, K., Murayama, E., Zapata, A., Cortes, A., Perret, E., Machu, C. and Herbomel, P. (2008) 'Live imaging of emerging hematopoietic stem cells and early thymus colonization', *Blood,* 111(3), pp. 1147-56.

Klepikova, A. V., Kasianov, A. S., Chesnokov, M. S., Lazarevich, N. L., Penin, A. A. and Logacheva, M. (2017) 'Effect of method of deduplication on estimation of differential gene expression using RNA-seq', *PeerJ*.

Kobayashi, I., Kobayashi-Sun, J., Kim, A. D., Pouget, C., Fujita, N., Suda, T. and Traver, D. (2014) 'Jam1a-Jam2a interactions regulate haematopoietic stem cell fate through Notch signalling', *Nature,* 512(7514), pp. 319-23.

Kondo, M., Weissman, I. L. and Akashi, K. (1997) 'Identification of clonogenic common lymphoid progenitors in mouse bone marrow', *Cell,* 91(5), pp. 661-72.

Kumaravelu, P., Hook, L., Morrison, A. M., Ure, J., Zhao, S., Zuyev, S., Ansell, J. and Medvinsky, A. (2002) 'Quantitative developmental anatomy of definitive haematopoietic stem cells/long-term repopulating units (HSC/RUs): role of the aorta-gonad-mesonephros (AGM) region and the yolk sac in colonisation of the mouse embryonic liver', *Development,* 129(21), pp. 4891-9.

Kwan, W., Cortes, M., Frost, I., Esain, V., Theodore, L. N., Liu, S. Y., Budrow, N., Goessling, W. and North, T. E. (2016) 'The Central Nervous System Regulates Embryonic HSPC Production via Stress-Responsive Glucocorticoid Receptor Signaling', *Cell Stem Cell,* 19(3), pp. 370-82.

Labastie, M. C., Cortes, F., Romeo, P. H., Dulac, C. and Peault, B. (1998) 'Molecular identity of hematopoietic precursor cells emerging in the human embryo', *Blood,* 92(10), pp. 3624-35.

Lan, Y., He, W. Y., Li, Z., Wang, Y., Wang, J., Gao, J., Wang, W. L., Cheng, T., Liu, B. and Yang, X. (2014) 'Endothelial Smad4 restrains the transition to hematopoietic progenitors via suppression of ERK activation', *Blood,* 123(14), pp. 2161-2171.

Lancrin, C., Sroczynska, P., Stephenson, C., Allen, T., Kouskoff, V. and Lacaud, G. (2009) 'The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage', *Nature,* 457(7231), pp. 892-5.

Laplante, M. and Sabatini, D. M. (2009) 'mTOR signaling at a glance'.

Lassila, O., Eskola, J., Toivanen, P., Martin, C. and Dieterlen-Lievre, F. (1978) 'The origin of lymphoid stem cells studied in chick yold sac-embryo chimaeras', *Nature,* 272(5651), pp. 353-4.

Lawson, N. D., Scheer, N., Pham, V. N., Kim, C. H., Chitnis, A. B., Campos-Ortega, J. A. and Weinstein, B. M. (2001) 'Notch signaling is required for arterial-venous differentiation during embryonic vascular development', *Development,* 128(19), pp. 3675-83.

Lee, H. O., Levorse, J. M. and Shin, M. K. (2003) 'The endothelin receptor-B is required for the migration of neural crest-derived melanocyte and enteric neuron precursors', *Dev Biol,* 259(1), pp. 162-75.

Lemischka, I. R., Raulet, D. H. and Mulligan, R. C. (1986) 'Developmental potential and dynamic behavior of hematopoietic stem cells', *Cell,* 45(6), pp. 917-27.

Levine, J. H., Simonds, E. F., Bendall, S. C., Davis, K. L., Amir el, A. D., Tadmor, M. D., Litvin, O., Fienberg, H. G., Jager, A., Zunder, E. R., Finck, R., Gedman, A. L., Radtke, I., Downing, J. R., Pe'er, D. and Nolan, G. P. (2015) 'Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis', *Cell,* 162(1), pp. 184-97.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics: Vol. 16*, pp. 2078-9.

Li, Y., Esain, V., Teng, L., Xu, J., Kwan, W., Frost, I. M., Yzaguirre, A. D., Cai, X., Cortes, M., Maijenburg, M. W., Tober, J., Dzierzak, E., Orkin, S. H., Tan, K., North, T. E. and Speck, N. A. (2014) 'Inflammatory signaling regulates embryonic hematopoietic stem and progenitor cell production', *Genes Dev,* 28(23), pp. 2597-612.

Li, Z., Lan, Y., He, W., Chen, D., Wang, J., Zhou, F., Wang, Y., Sun, H., Chen, X., Xu, C., Li, S., Pang, Y., Zhang, G., Yang, L., Zhu, L., Fan, M., Shang, A., Ju, Z., Luo, L., Ding, Y., Guo, W., Yuan, W., Yang, X. and Liu, B. (2012) 'Mouse embryonic head as a site for hematopoietic stem cell development', *Cell Stem Cell,* 11(5), pp. 663-75.

Li, Z., Vink, C. S., Mariani, S. A. and Dzierzak, E. (2016) 'Subregional localization and characterization of Ly6aGFP-expressing hematopoietic cells in the mouse embryonic head', *Dev Biol,* 416(1), pp. 34-41.

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P. and Tamayo, P. (2015) 'The Molecular Signatures Database (MSigDB) hallmark gene set collection', *Cell Syst,* 1(6), pp. 417-425.

Lichtinger, M., Ingram, R., Hannah, R., Muller, D., Clarke, D., Assi, S. A., Lie, A. L. M., Noailles, L., Vijayabaskar, M. S., Wu, M., Tenen, D. G., Westhead, D. R., Kouskoff, V., Lacaud, G., Gottgens, B. and Bonifer, C. (2012) 'RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis', *Embo j,* 31(22), pp. 4318-33.

Lie, A. L. M., Marinopoulou, E., Li, Y., Patel, R., Stefanska, M., Bonifer, C., Miller, C., Kouskoff, V. and Lacaud, G. (2014) 'RUNX1 positively regulates a cell adhesion and migration program in murine hemogenic endothelium prior to blood emergence', *Blood,* 124(11), pp. e11-20.

Lin, C. T., Datta, V., Okwan-Duodu, D., Chen, X., Fuchs, S., Alsabeh, R., Billet, S., Bernstein, K. E. and Shen, X. Z. (2011) 'Angiotensin-converting enzyme is required for normal myelopoiesis', *Faseb Journal,* 25(4), pp. 1145-1155.

Lin, H., Sangmal, M., Smith, M. J., Jr. and Young, D. B. (1993) 'Effect of endothelin-1 on glomerular hydraulic pressure and renin release in dogs', *Hypertension,* 21(6 Pt 1), pp. 845-51.

Lis, R., Karrasch, C. C., Poulos, M. G., Kunar, B., Redmond, D., Duran, J. G. B., Badwe, C. R., Schachterle, W., Ginsberg, M., Xiang, J., Tabrizi, A. R., Shido, K., Rosenwaks, Z., Elemento, O., Speck, N. A., Butler, J. M., Scandura, J. M. and Rafii, S. (2017) 'Conversion of adult endothelium to immunocompetent haematopoietic stem cells', *Nature,* 545(7655), pp. 439.

Liu, F., Walmsley, M., Rodaway, A. and Patient, R. (2008) 'Fli1 acts at the top of the transcriptional network driving blood and endothelial development', *Curr Biol,* 18(16), pp. 1234-40.

Liu, Z., Tu, H., Kang, Y., Xue, Y., Ma, D., Zhao, C., Li, H., Wang, L. and Liu, F. (2019) 'Primary cilia regulate hematopoietic stem and progenitor cell specification through Notch signaling in zebrafish', *Nature Communications,* 10(1), pp. 1839.

Lizama, C. O., Hawkins, J. S., Schmitt, C. E., Bos, F. L., Zape, J. P., Cautivo, K. M., Borges Pinto, H., Rhyner, A. M., Yu, H., Donohoe, M. E., Wythe, J. D. and Zovein, A. C. (2015) 'Repression of arterial genes in hemogenic endothelium is sufficient for haematopoietic fate acquisition', *Nat Commun,* 6, pp. 7739.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biol: Vol. 12*.

Luckett, W. P. (1978) 'Origin and differentiation of the yolk sac and extraembryonic mesoderm in presomite human and rhesus monkey embryos', *Am J Anat,* 152(1), pp. 59-97.

Maeno, M., Tochinai, S. and Katagiri, C. (1985) 'Differential participation of ventral and dorsolateral mesoderms in the hemopoiesis of Xenopus, as revealed in diploid-triploid or interspecific chimeras', *Dev Biol,* 110(2), pp. 503-8.

Majhail, N. S. (2017) 'Long Term Complications After Hematopoietic Cell Transplantation', *Hematol Oncol Stem Cell Ther,* 10(4), pp. 220-7.

Manno, G. L., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnerberg, P., Furlan, A., Fan, J., Borm, L. E., Liu, Z., Bruggen, D. v., Guo, J., He, X., Barker, R., Sundström, E., Castelo-Branco, G., Cramer, P., Adameyko, I., Linnarsson, S. and Kharchenko, P. V. (2018) 'RNA velocity of single cells', *Nature,* 560(7719), pp. 494.

Mansson, R., Hultquist, A., Luc, S., Yang, L., Anderson, K., Kharazi, S., Al-Hashmi, S., Liuba, K., Thoren, L., Adolfsson, J., Buza-Vidas, N., Qian, H., Soneji, S., Enver, T., Sigvardsson, M. and Jacobsen, S. E. (2007) 'Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors', *Immunity,* 26(4), pp. 407-19.

Mariani, S. A., Li, Z., Rice, S., Krieg, C., Fragkogianni, S., Robinson, M., Vink, C. S., Pollard, J. W. and Dzierzak, E. (2019) 'Pro-inflammatory Aorta-Associated Macrophages Are Involved in Embryonic Development of Hematopoietic Stem Cells', *Immunity,* 50(6), pp. 1439-1452.e5.

Mascarenhas, M. I., Parker, A., Dzierzak, E. and Ottersbach, K. (2009) 'Identification of novel regulators of hematopoietic stem cell development through refinement of stem cell localization and expression profiling', *Blood,* 114(21), pp. 4645-53.

McGarvey, A. C., Rybtsov, S., Souilhol, C., Tamagno, S., Rice, R., Hills, D., Godwin, D., Rice, D., Tomlinson, S. R. and Medvinsky, A. (2017) 'A molecular roadmap of the AGM region reveals BMPER as a novel regulator of HSC maturation', *The Journal of Experimental Medicine,* 214(12), pp. 3731-3751.

McGrath, K. E., Frame, J. M., Fromm, G. J., Koniski, A. D., Kingsley, P. D., Little, J., Bulger, M. and Palis, J. (2011) 'A transient definitive erythroid lineage with unique regulation of the beta-globin locus in the mammalian embryo', *Blood,* 117(17), pp. 4600-8.

McInnes, L., Healy, J. and Melville, J. (2018) 'UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction'.

Medvinsky, A. and Dzierzak, E. (1996) 'Definitive hematopoiesis is autonomously initiated by the AGM region', *Cell,* 86(6), pp. 897-906.

Medvinsky, A., Rybtsov, S. and Taoudi, S. (2011) 'Embryonic origin of the adult hematopoietic system: advances and questions'.

Medvinsky, A., Taoudi, S., Mendes, S. and Dzierzak, E. (2008) 'Analysis and manipulation of hematopoietic progenitor and stem cells from murine embryonic tissues', *Curr Protoc Stem Cell Biol,* Chapter 2, pp. Unit 2A.6.

Medvinsky, A. L., Gan, O. I., Semenova, M. L. and Samoylina, N. L. (1996) 'Development of day-8 colony-forming unit-spleen hematopoietic progenitors during early murine embryogenesis: spatial and temporal mapping', *Blood,* 87(2), pp. 557-66.

Medvinsky, A. L., Samoylina, N. L., Muller, A. M. and Dzierzak, E. A. (1993) 'An early pre-liver intraembryonic source of CFU-S in the developing mouse', *Nature,* 364(6432), pp. 64-7.

Messerli, F. H., Bangalore, S., Bavishi, C. and Rimoldi, S. F. (2018) 'Angiotensin-Converting Enzyme Inhibitors in Hypertension To Use or Not to Use?', *Journal of the American College of Cardiology,* 71(13), pp. 1474-1482.

Miller, J. D., Clabaugh, S. E., Smith, D. R., Stevens, R. B. and Wrenshall, L. E. (2012) 'Interleukin-2 is present in human blood vessels and released in biologically active form by heparanase', *Immunology and Cell Biology,* 90(2), pp. 159-167.

Molin, D. G. M., DeRuiter, M. C., Wisse, L. J., Azhar, M., Doetschman, T., Poelmann, R. E. and Gittenberger-de Groot, A. C. (2002) 'Altered apoptosis pattern during pharyngeal arch artery remodelling is associated with aortic arch malformations in Tgf beta 2 knock-out mice', *Cardiovascular Research,* 56(2), pp. 312-322.

Monteiro, R., Pinheiro, P., Joseph, N., Peterkin, T., Koth, J., Repapi, E., Bonkhofer, F., Kirmizitas, A. and Patient, R. (2016) 'Transforming Growth Factor beta Drives Hemogenic Endothelium Programming and the Transition to Hematopoietic Stem Cells', *Developmental Cell,* 38(4), pp. 358-370.

Mousavi, S. H., Abroun, S., Soleimani, M. and Mowla, S. J. (2015) 'Expansion of human cord blood hematopoietic stem/progenitor cells in three-dimensional Nanoscaffold coated with Fibronectin', *Int J Hematol Oncol Stem Cell Res: Vol. 2*, pp. 72-9.

Muller, A. M., Medvinsky, A., Strouboulis, J., Grosveld, F. and Dzierzak, E. (1994) 'Development of hematopoietic stem cell activity in the mouse embryo', *Immunity,* 1(4), pp. 291-301.

Ng, E. S., Azzola, L., Bruveris, F. F., Calvanese, V., Phipson, B., Vlahos, K., Hirst, C., Jokubaitis, V. J., Yu, Q. C., Maksimovic, J., Liebscher, S., Januar, V., Zhang, Z., Williams, B., Conscience, A., Durnall, J., Jackson, S., Costa, M., Elliott, D., Haylock, D. N., Nilsson, S. K., Saffery, R., Schenke-Layland, K., Oshlack, A., Mikkola, H. K., Stanley, E. G. and Elefanty, A. G. (2016) 'Differentiation of human embryonic stem cells to HOXA(+) hemogenic vasculature that resembles the aorta-gonad-mesonephros', *Nat Biotechnol,* 34(11), pp. 1168-1179.

Nichterwitz, S., Chen, G., Aguila Benitez, J., Yilmaz, M., Storvall, H., Cao, M., Sandberg, R., Deng, Q. and Hedlund, E. (2016) 'Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling', *Nat Commun,* 7, pp. 12139.

Nilsson, S. K., Johnston, H. M., Whitty, G. A., Williams, B., Webb, R. J., Denhardt, D. T., Bertoncello, I., Bendall, L. J., Simmons, P. J. and Haylock, D. N. (2005) 'Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells', *Blood,* 106(4), pp. 1232-9.

Nishimura, D. (2004) 'BioCarta', *https://home.liebertpub.com/bsi*.

North, T., Gu, T. L., Stacy, T., Wang, Q., Howard, L., Binder, M., Marin-Padilla, M. and Speck, N. A. (1999) 'Cbfa2 is required for the formation of intra-aortic hematopoietic clusters', *Development,* 126(11), pp. 2563-75.

North, T. E., Goessling, W., Peeters, M., Li, P., Ceol, C., Lord, A. M., Weber, G. J., Harris, J., Cutting, C. C., Huang, P., Dzierzak, E. and Zon, L. I. (2009) 'Hematopoietic stem cell development is dependent on blood flow', *Cell,* 137(4), pp. 736-48.

North, T. E., Goessling, W., Walkley, C. R., Lengerke, C., Kopani, K. R., Lord, A. M., Weber, G. J., Bowman, T. V., Jang, I.-H., Grosser, T., FitzGerald, G. A., Daley, G. Q., Orkin, S. H. and Zon, L. I. (2007) 'Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis', *Nature,* 447(7147), pp. 1007.

Nostro, M. C., Cheng, X., Keller, G. M. and Gadue, P. (2008) 'Wnt, activin, and BMP signaling regulate distinct stages in the developmental pathway from embryonic stem cells to blood', *Cell Stem Cell,* 2(1), pp. 60-71.

Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O. I., Wilson, G., Kaufmann, K. B., McLeod, J., Laurenti, E., Dunant, C. F., McPherson, J. D., Stein, L. D., Dror, Y. and Dick, J. E. (2016) 'Distinct routes of lineage development reshape the human blood hierarchy across ontogeny'.

Nottingham, W. T., Jarratt, A., Burgess, M., Speck, C. L., Cheng, J. F., Prabhakar, S., Rubin, E. M., Li, P. S., Sloane-Stanley, J., Kong, A. S. J. and de Bruijn, M. F. (2007) 'Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer', *Blood,* 110(13), pp. 4188-97.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W., Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O'Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D., Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D. and Pruitt, K. D. (2016) 'Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation', *Nucleic Acids Res,* 44(D1), pp. D733-45.

Oatley, M., Bölükbası, Ö. V., Svensson, V., Shvartsman, M., Ganter, K., Zirngibl, K., Pavlovich, P. V., Milchevskaya, V., Foteva, V., Natarajan, K. N., Baying, B., Benes, V., Patil, K. R., Teichmann, S. A. and Lancrin, C. (2020) 'Single-cell transcriptomics identifies CD44 as a marker and regulator of endothelial to haematopoietic transition', *Nature Communications,* 11(1), pp. 1-18.

Oberlin, E., Tavian, M., Blazsek, I. and Peault, B. (2002) 'Blood-forming potential of vascular endothelium in the human embryo', *Development,* 129(17), pp. 4147-57.

Olsson, A., Venkatasubramanian, M., Chaudhri, V. K., Aronow, B. J., Salomonis, N., Singh, H. and Grimes, H. L. (2016) 'Single-cell analysis of mixed-lineage states leading to a binary cell fate choice', *Nature,* 537(7622), pp. 698-702.

Ott, M., Amunts, A. and Brown, A. (2016) 'Organization and Regulation of Mitochondrial Protein Synthesis*', *http://dx.doi.org/10.1146/annurev-biochem-060815-014334*.

Ottersbach, K. and Dzierzak, E. (2005) 'The murine placenta contains hematopoietic stem cells within the vascular labyrinth region', *Dev Cell,* 8(3), pp. 377-87.

Palis, J., Robertson, S., Kennedy, M., Wall, C. and Keller, G. (1999) 'Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse', *Development,* 126(22), pp. 5073-84.

Pardanaud, L., Luton, D., Prigent, M., Bourcheix, L. M., Catala, M. and Dieterlen-Lievre, F. (1996) 'Two distinct endothelial lineages in ontogeny, one of them related to hemopoiesis', *Development,* 122(5), pp. 1363-71.

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. and Hellmann, I. (2016) 'The impact of amplification on differential expression analyses by RNA-seq', *Scientific Reports,* 6.

Paul, F., Arkin, Y., Giladi, A., Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., David, E., Cohen, N., Lauridsen, F. K., Haas, S., Schlitzer, A., Mildner, A., Ginhoux, F., Jung, S., Trumpp, A., Porse, B. T., Tanay, A. and Amit, I. (2015) 'Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors', *Cell,* 163(7), pp. 1663-77.

Peeters, M., Ottersbach, K., Bollerot, K., Orelio, C., de Bruijn, M., Wijgerde, M. and Dzierzak, E. (2009) 'Ventral embryonic tissues and Hedgehog proteins induce early AGM hematopoietic stem cell development', *Development,* 136(15), pp. 2613-21.

Pereira, C., Clarke, E. and Damen, J. (2007) 'Hematopoietic colony-forming cell assays', *Methods Mol Biol,* 407, pp. 177-208.

Pereira, C. F., Chang, B., Qiu, J., Niu, X., Papatsenko, D., Hendry, C. E., Clark, N. R., Nomura-Kitabayashi, A., Kovacic, J. C., Ma'ayan, A., Schaniel, C., Lemischka, I. R. and Moore, K. (2013) 'Induction of a hemogenic program in mouse fibroblasts', *Cell Stem Cell,* 13(2), pp. 205-18.

Picelli, S., Faridani, O. R., Bjorklund, A. K., Winberg, G., Sagasser, S. and Sandberg, R. (2014) 'Full-length RNA-seq from single cells using Smart-seq2', *Nature Protocols,* 9(1), pp. 171-181.

Pick, M., Azzola, L., Mossman, A., Stanley, E. G. and Elefanty, A. G. (2007) 'Differentiation of human embryonic stem cells in serum-free medium reveals distinct roles for bone morphogenetic protein 4, vascular endothelial growth factor, stem cell factor, and fibroblast growth factor 2 in hematopoiesis', *Stem Cells,* 25(9), pp. 2206-14.

Plein, A., Fantin, A., Denti, L., Pollard, J. W. and Ruhrberg, C. (2018) 'Erythro-myeloid progenitors contribute endothelial cells to blood vessels', *Nature,* 562(7726), pp. 223-8.

Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics,* 26(6), pp. 841-2.

Rhodes, K. E., Gekas, C., Wang, Y., Lux, C. T., Francis, C. S., Chan, D. N., Conway, S., Orkin, S. H., Yoder, M. C. and Mikkola, H. K. (2008) 'The emergence of hematopoietic stem cells is initiated in the placental vasculature in the absence of circulation', *Cell Stem Cell,* 2(3), pp. 252-63.

Richmond, R. S., Tallant, E. A., Gallagher, P. E., Ferrario, C. M. and Strawn, W. B. (2004) 'Angiotensin II stimulates arachidonic acid release from bone marrow stromal cells', *Journal of the Renin-Angiotensin-Aldosterone System,* 5(4), pp. 176-182.

Robin, C., Ottersbach, K., Durand, C., Peeters, M., Vanes, L., Tybulewicz, V. and Dzierzak, E. (2006) 'An unexpected role for IL-3 in the embryonic development of hematopoietic stem cells', *Dev Cell,* 11(2), pp. 171-80.

Ronn, R. E., Guibentif, C., Moraghebi, R., Chaves, P., Saxena, S., Garcia, B. and Woods, N. B. (2015) 'Retinoic acid regulates hematopoietic development from human pluripotent stem cells', *Stem Cell Reports,* 4(2), pp. 269-81.

Rossi, G. P., Albertin, G., Bova, S., Belloni, A. S., Fallo, F., Pagotto, U., Trevisi, L., Palu, G., Pessina, A. C. and Nussdorfer, G. G. (1997) 'Autocrine-paracrine role of endothelin-1 in the regulation of aldosterone synthase expression and intracellular Ca2+ in human adrenocortical carcinoma NCI-H295 cells', *Endocrinology,* 138(10), pp. 4421-6.

Ruiz-Herguido, C., Guiu, J., D'Altri, T., Ingles-Esteve, J., Dzierzak, E., Espinosa, L. and Bigas, A. (2012) 'Hematopoietic stem cell development requires transient Wnt/beta-catenin activity', *J Exp Med,* 209(8), pp. 1457-68.

Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O. and Medvinsky, A. (2014) 'Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43(-) embryonic precursor', *Stem Cell Reports,* 3(3), pp. 489-501.

Rybtsov, S., Ivanovs, A., Zhao, S. and Medvinsky, A. (2016) 'Concealed expansion of immature precursors underpins acute burst of adult HSC activity in foetal liver'.

Rybtsov, S., Sobiesiak, M., Taoudi, S., Souilhol, C., Senserrich, J., Liakhovitskaia, A., Ivanovs, A., Frampton, J., Zhao, S. L. and Medvinsky, A. (2011) 'Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region', *Journal of Experimental Medicine,* 208(6), pp. 1305-1315.

Sagara, S., Sugaya, K., Tokoro, Y., Tanaka, S., Takano, H., Kodama, H., Nakauchi, H. and Takahama, Y. (1997) 'B220 expression by T lymphoid progenitor cells in mouse fetal liver', *J Immunol,* 158(2), pp. 666-76.

Sandler, V. M., Lis, R., Liu, Y., Kedem, A., James, D., Elemento, O., Butler, J. M., Scandura, J. M. and Rafii, S. (2014) 'Reprogramming human endothelial cells to haematopoietic cells requires vascular induction', *Nature,* 511(7509), pp. 312-8.

Sanjuan-Pla, A., Macaulay, I. C., Jensen, C. T., Woll, P. S., Luis, T. C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Jones, T. B., Chowdhury, O., Stenson, L., Lutteropp, M., Green, J. C. A., Facchini, R., Boukarabila, H., Grover, A., Gambardella, A., Thongjuea, S., Carrelha, J., Tarrant, P., Atkinson, D., Clark, S.-A., Nerlov, C. and Jacobsen, S. E. W. (2013) 'Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy', *Nature,* 502(7470), pp. 232.

Savary, K., Michaud, A., Favier, J., Larger, E., Corvol, P. and Gasc, J. M. (2005) 'Role of the renin-angiotensin system in primitive erythropoiesis in the chick embryo', *Blood,* 105(1), pp. 103-110.

Sawamiphak, S., Kontarakis, Z. and Stainier, D. Y. (2014) 'Interferon gamma signaling positively regulates hematopoietic stem cell emergence', *Dev Cell,* 31(5), pp. 640-53.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. and Cardona, A. (2012) 'Fiji: an open-source platform for biological-image analysis', *Nat Methods,* 9(7), pp. 676-82.

Shalaby, F., Rossant, J., Yamaguchi, T. P., Gertsenstein, M., Wu, X. F., Breitman, M. L. and Schuh, A. C. (1995) 'Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice', *Nature,* 376(6535), pp. 62-6.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) 'Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks', *Genome Res: Vol. 11*, pp. 2498-504.

Shioide, M. and Noda, M. (1993) 'Endothelin modulates osteopontin and osteocalcin messenger ribonucleic acid expression in rat osteoblastic osteosarcoma cells', *J Cell Biochem,* 53(2), pp. 176-80.

Sinka, L., Biasch, K., Khazaal, I., Peault, B. and Tavian, M. (2012) 'Angiotensin-converting enzyme (CD143) specifies emerging lympho-hematopoietic progenitors in the human embryo', *Blood,* 119(16), pp. 3712-3723.

Sitnicka, E., Ruscetti, F. W., Priestley, G. V., Wolf, N. S. and Bartelmez, S. H. (1996) 'Transforming growth factor beta(1) directly and reversibly inhibits the initial cell divisions of long-term repopulating hematopoietic stem cells', *Blood,* 88(1), pp. 82-88.

Slukvin, II (2013) 'Deciphering the hierarchy of angiohematopoietic progenitors from human pluripotent stem cells', *Cell Cycle,* 12(5), pp. 720-7.

Souilhol, C., Gonneau, C., Lendinez, J. G., Batsivari, A., Rybtsov, S., Wilson, H., Morgado-Palacin, L., Hills, D., Taoudi, S., Antonchuk, J., Zhao, S. L. and Medvinsky, A. (2016a) 'Inductive interactions mediated by interplay of asymmetric signalling underlie development of adult haematopoietic stem cells', *Nature Communications,* 7, pp. 13.

Souilhol, C., Lendinez, J. G., Rybtsov, S., Murphy, F., Wilson, H., Hills, D., Batsivari, A., Binagui-Casas, A., McGarvey, A. C., MacDonald, H. R., Kageyama, R., Siebel, C., Zhao, S. L. and Medvinsky, A. (2016b) 'Developing HSCs become Notch independent by the end of maturation in the AGM region', *Blood,* 128(12), pp. 1567-1577.

Spangrude, G. J., Heimfeld, S. and Weissman, I. L. (1988) 'Purification and characterization of mouse hematopoietic stem cells', *Science,* 241(4861), pp. 58-62.

Sparks, M. A., Crowley, S. D., Gurley, S. B., Mirotsou, M. and Coffman, T. M. (2014) 'Classical Renin-Angiotensin System in Kidney Physiology', *Comprehensive Physiology,* 4(3), pp. 1201-1228.

Sturgeon, C. M., Ditadi, A., Awong, G., Kennedy, M. and Keller, G. (2014) 'Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells', *Nat Biotechnol,* 32(6), pp. 554-61.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. and Mesirov, J. P. (2005) 'Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles', *Proc Natl Acad Sci U S A,* 102(43), pp. 15545-50.

Sugimura, R., Jha, D. K., Han, A., Soria-Valles, C., da Rocha, E. L., Lu, Y. F., Goettel, J. A., Serrao, E., Rowe, R. G., Malleshaiah, M., Wong, I., Sousa, P., Zhu, T. N., Ditadi, A., Keller, G., Engelman, A. N., Snapper, S. B., Doulatov, S. and Daley, G. Q. (2017) 'Haematopoietic stem and progenitor cells from human pluripotent stem cells', *Nature,* 545(7655), pp. 432-8.

Sung, C. P., Arleth, A. J., Storer, B. L. and Ohlstein, E. H. (1994) 'ANGIOTENSIN TYPE-1 RECEPTORS MEDIATE SMOOTH-MUSCLE PROLIFERATION AND ENDOTHELIN BIOSYNTHESIS IN RAT VASCULAR SMOOTH-MUSCLE', *Journal of Pharmacology and Experimental Therapeutics,* 271(1), pp. 429-437.

Szilvassy, S. J., Humphries, R. K., Lansdorp, P. M., Eaves, A. C. and Eaves, C. J. (1990) 'Quantitative assay for totipotent reconstituting hematopoietic stem cells by a competitive repopulation strategy', *Proc Natl Acad Sci U S A,* 87(22), pp. 8736-40.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J. and von Mering, C. (2015) 'STRING v10: protein–protein interaction networks, integrated over the tree of life', *Nucleic Acids Res,* 43(Database issue), pp. D447-52.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J. and Mering, C. V. (2019) 'STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets', *Nucleic Acids Res,* 47(D1), pp. D607-d613.

Tamplin, O. J., Durand, E. M., Carr, L. A., Childs, S. J., Hagedorn, E. J., Li, P., Yzaguirre, A. D., Speck, N. A. and Zon, L. I. (2015) 'Hematopoietic stem cell arrival triggers dynamic remodeling of the perivascular niche', *Cell,* 160(1-2), pp. 241-52.

Tang, Y., Bai, H., Urs, S., Wang, Z. and Liaw, L. (2013) 'Notch1 activation in embryonic VE-cadherin populations selectively blocks hematopoietic stem cell generation and fetal liver hematopoiesis', *Transgenic Res,* 22(2), pp. 403-10.

Taoudi, S., Gonneau, C., Moore, K., Sheridan, J. M., Blackburn, C. C., Taylor, E. and Medvinsky, A. (2008) 'Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-cadherin+CD45+ pre-definitive HSCs', *Cell Stem Cell,* 3(1), pp. 99-108.

Taoudi, S. and Medvinsky, A. (2007a) 'Functional identification of the hematopoietic stem cell niche in the ventral domain of the embryonic dorsal aorta'.

Taoudi, S. and Medvinsky, A. (2007b) 'Functional identification of the hematopoietic stem cell niche in the ventral domain of the embryonic dorsal aorta', *Proceedings of the National Academy of Sciences of the United States of America,* 104(22), pp. 9399-9403.

Taoudi, S., Morrison, A. M., Inoue, H., Gribi, R., Ure, J. and Medvinsky, A. (2005) 'Progressive divergence of definitive haematopoietic stem cells from the endothelial compartment does not depend on contact with the foetal liver', *Development,* 132(18), pp. 4179-91.

Tavian, M., Coulombel, L., Luton, D., SanClemente, H., DieterlenLievre, F. and Peault, B. (1996) 'Aorta-associated CD34(+) hematopoietic cells in the early human embryo', *Blood,* 87(1), pp. 67-72.

Tavian, M., Hallais, M. F. and Peault, B. (1999) 'Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo', *Development,* 126(4), pp. 793-803.

Taylor, E., Taoudi, S. and Medvinsky, A. (2010) 'Hematopoietic stem cell activity in the aorta-gonad-mesonephros region enhances after mid-day 11 of mouse development', *Int J Dev Biol,* 54(6-7), pp. 1055-60.

'The Gene Ontology Resource: 20 years and still GOing strong', (2019) *Nucleic Acids Res,* 47(D1), pp. D330-d338.

Tober, J., Koniski, A., McGrath, K. E., Vemishetti, R., Emerson, R., de Mesy-Bentley, K. K., Waugh, R. and Palis, J. (2007) 'The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis', *Blood,* 109(4), pp. 1433-41.

Travnickova, J., Tran Chau, V., Julien, E., Mateos-Langerak, J., Gonzalez, C., Lelievre, E., Lutfalla, G., Tavian, M. and Kissa, K. (2015) 'Primitive macrophages control HSPC mobilization and definitive haematopoiesis', *Nat Commun,* 6, pp. 6227.

Turpen, J. B., Knudson, C. M. and Hoefen, P. S. (1981) 'The early ontogeny of hematopoietic cells studied by grafting cytogenetically labeled tissue anlagen: localization of a prospective stem cell compartment', *Dev Biol,* 85(1), pp. 99-112.

Velten, L., Haas, S. F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B. P., Hirche, C., Lutz, C., Buss, E. C., Nowak, D., Boch, T., Hofmann, W. K., Ho, A. D., Huber, W., Trumpp, A., Essers, M. A. and Steinmetz, L. M. (2017) 'Human haematopoietic stem cell lineage commitment is a continuous process', *Nat Cell Biol,* 19(4), pp. 271-281.

Vining, K. H. and Mooney, D. J. (2017) 'Mechanical forces direct stem cell behaviour in development and regeneration', *Nat Rev Mol Cell Biol,* 18(12), pp. 728-42.

Wahlster, L. and Daley, G. Q. (2016) 'Progress towards generation of human haematopoietic stem cells', *Nature Cell Biology,* 18(11), pp. 1111.

Wang, L., Zhang, P., Wei, Y., Gao, Y., Patient, R. and Liu, F. (2011) 'A blood flow-dependent klf2a-NO signaling cascade is required for stabilization of hematopoietic stem cell programming in zebrafish embryos', *Blood,* 118(15), pp. 4102-10.

Wang, Y. and Nakayama, N. (2009) 'WNT and BMP signaling are both required for hematopoietic cell development from human ES cells', *Stem Cell Res,* 3(2-3), pp. 113-25.

Watford, W. T., Hissong, B. D., Bream, J. H., Kanno, Y., Muul, L. and O'Shea, J. J. (2004) 'Signaling by IL-12 and IL-23 and the immunoregulatory roles of STAT4', *Immunol Rev,* 202, pp. 139-56.

Watt, S. M., Butler, L. H., Tavian, M., Buhring, H. J., Rappold, I., Simmons, P. J., Zannettino, A. C., Buck, D., Fuchs, A., Doyonnas, R., Chan, J. Y., Levesque, J. P., Peault, B. and Roxanis, I. (2000) 'Functionally defined CD164 epitopes are expressed on CD34(+) cells throughout ontogeny but display distinct distribution patterns in adult hematopoietic and nonhematopoietic tissues', *Blood,* 95(10), pp. 3113-24.

Weijts, B., Gutierrez, E., Saikin, S. K., Ablooglu, A. J., Traver, D., Groisman, A. and Tkachenko, E. (2018) 'Blood flow-induced Notch activation and endothelial migration enable vascular remodeling in zebrafish embryos', *Nature Communications,* 9(1), pp. 5314.

Wickham, H. (2019) *ggplot2 - Elegant Graphics for Data Analysis.* Springer.

Wilkinson, A. C., Ishida, R., Kikuchi, M., Sudo, K., Morita, M., Crisostomo, R. V., Yamamoto, R., Loh, K. M., Nakamura, Y., Watanabe, M., Nakauchi, H. and Yamazaki, S. (2019) 'Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation', *Nature,* 571(7763), pp. 117-121.

Wilkinson, R. N., Pouget, C., Gering, M., Russell, A. J., Davies, S. G., Kimelman, D. and Patient, R. (2009) 'Hedgehog and Bmp polarize hematopoietic stem cell emergence in the zebrafish dorsal aorta', *Dev Cell,* 16(6), pp. 909-16.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R. C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C. F., Eshkind, L., Bockamp, E., Lio, P., MacDonald, H. R. and Trumpp, A. (2008) 'Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair', *Cell,* 135(6), pp. 1118-1129.

Wolf, F. A., Angerer, P. and Theis, F. J. (2018) 'SCANPY: large-scale single-cell gene expression data analysis', *Genome Biol,* 19(1), pp. 15.

Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Gottgens, B., Rajewsky, N., Simon, L. and Theis, F. J. (2019) 'PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells', *Genome Biol,* 20(1), pp. 59.

Yamada, H., Tsubakimoto, Y., Yokoi, H., Kishida, S., Kato, T., Kawahito, H., Hirai, H., Asihara, E., Maekawa, T., Ikeda, K., Takahashi, T., Okigaki, M. and Matsubara, H. (2010) 'Bone marrow angiotensin AT1 receptor regulates differentiation of monocyte lineage progenitors from hematopoietic stem cells', *Endocrine Journal,* 57, pp. S607-S607.

Yamazaki, S., Ema, H., Karlsson, G., Yamaguchi, T., Miyoshi, H., Shioda, S., Taketo, M. M., Karlsson, S., Iwama, A. and Nakauchi, H. (2011) 'Nonmyelinating Schwann Cells Maintain Hematopoietic Stem Cell Hibernation in the Bone Marrow Niche', *Cell,* 147(5), pp. 1146-1158.

Yang, Q. W., Liu, X. L., Zhou, T., Cook, J., Nguyen, K. and Bai, X. Y. (2016) 'RNA polymerase II pausing modulates hematopoietic stem cell emergence in zebrafish', *Blood,* 128(13), pp. 1701-1710.

Yoder, M. C. (2014) 'Inducing definitive hematopoiesis in a dish', *Nature Biotechnology,* 32(6), pp. 539-541.

Yokomizo, T. and Dzierzak, E. (2010) 'Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos'.

Yokomizo, T., Ogawa, M., Osato, M., Kanno, T., Yoshida, H., Fujimoto, T., Fraser, S., Nishikawa, S., Okada, H., Satake, M., Noda, T. and Ito, Y. (2001) 'Requirement of Runx1/AML1/PEBP2alphaB for the generation of haematopoietic cells from endothelial cells', *Genes Cells,* 6(1), pp. 13-23.

Yokomizo, T., Watanabe, N., Umemoto, T., Matsuo, J., Harai, R., Kihara, Y., Nakamura, E., Tada, N., Sato, T., Takaku, T., Shimono, A., Takizawa, H., Nakagata, N., Mori, S., Kurokawa, M., Tenen, D. G., Osato, M., Suda, T. and Komatsu, N. (2019) 'Hlf marks the developmental pathway for hematopoietic stem cells but not for erythro-myeloid progenitors', *J Exp Med,* 216(7), pp. 1599-1614.

Yoshimoto, M., Porayette, P., Glosson, N. L., Conway, S. J., Carlesso, N., Cardoso, A. A., Kaplan, M. H. and Yoder, M. C. (2012) 'Autonomous murine T-cell progenitor production in the extra-embryonic yolk sac before HSC emergence', *Blood,* 119(24), pp. 5706-5714.

Yousufzai, S. Y. K. and AbdelLatif, A. A. (1997) 'Endothelin-1 stimulates the release of arachidonic acid and prostaglandins in cultured human ciliary muscle cells: Activation of phospholipase A(2)', *Experimental Eye Research,* 65(1), pp. 73-81.

Yuan, W., Qian, M., Li, Z. X., Zhao, C. L., Zhao, J. and Xiao, J. R. (2019) 'Endothelin-1 Activates the Notch Signaling Pathway and Promotes Tumorigenesis in Giant Cell Tumor of the Spine', *Spine (Phila Pa 1976)*.

Zambidis, E. T., Park, T. S., Yu, W., Tam, A., Levine, M., Yuan, X., Pryzhkova, M. and Peault, B. (2008) 'Expression of angiotensin-converting enzyme (CD143) identifies and regulates primitive hemangioblasts derived from human pluripotent stem cells', *Blood,* 112(9), pp. 3601-3614.

Zeng, Y., He, J., Bai, Z., Li, Z., Gong, Y., Liu, C., Ni, Y., Du, J., Ma, C., Bian, L., Lan, Y. and Liu, B. (2019) 'Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing', *Cell Research,* 29(11), pp. 881-894.

Zhang, C. X., Chen, Y. S., Sun, B. F., Wang, L., Yang, Y., Ma, D. Y., Lv, J. H., Heng, J., Ding, Y. Y., Xue, Y. Y., Lu, X. Y., Xiao, W., Yang, Y. G. and Liu, F. (2017) 'm(6)A modulates haematopoietic stem and progenitor cell specification', *Nature,* 549(7671), pp. 273-+.

Zhang, C. Y., Yin, H. M., Wang, H., Su, D., Xia, Y., Yan, L. F., Fang, B., Liu, W., Wang, Y. M., Gu, A. H. and Zhou, Y. (2018) 'Transforming growth factor-beta 1 regulates the nascent hematopoietic stem cell niche by promoting gluconeogenesis', *Leukemia,* 32(2), pp. 479-491.

Zhang, Y. Y., Clay, D., Mitjavila-Garcia, M. T., Alama, A., Mennesson, B., Berseneff, H., Louache, F., Bennaceur-Griscelli, A. and Oberlin, E. (2019) 'VE-Cadherin and ACE Co-Expression Marks Highly Proliferative Hematopoietic Stem Cells in Human Embryonic Liver', *Stem Cells and Development,* 28(3), pp. 165-185.

Zhao, M., Perry, J. M., Marshall, H., Venkatraman, A., Qian, P. X., He, X. C., Ahamed, J. and Li, L. H. (2014) 'Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells', *Nature Medicine,* 20(11), pp. 1321-1326.

Zhen, F., Lan, Y., Yan, B., Zhang, W. and Wen, Z. (2013) 'Hemogenic endothelium specification and hematopoietic stem cell maintenance employ distinct Scl isoforms'.

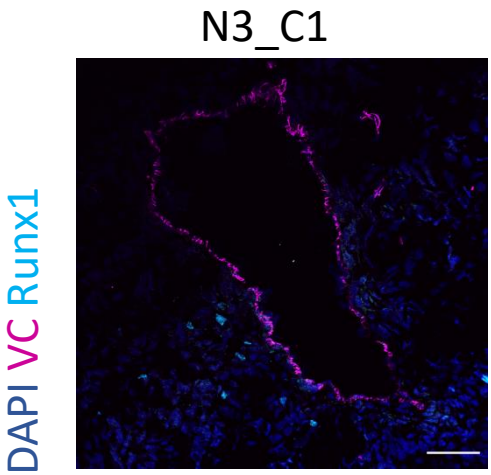
Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Zheng, X., Li, Z., Ni, Y., Mu, X., Wen, L., Cheng, T., Lan, Y., Yuan, W., Tang, F. and Liu, B. (2016) 'Tracing haematopoietic stem cell formation at single-cell resolution', *Nature,* 533(7604), pp. 487.

Zhou, J., Xu, J., Zhang, L., Liu, S., Ma, Y., Wen, X., Hao, J., Li, Z., Ni, Y., Li, X., Zhou, F., Li, Q., Wang, F., Wang, X., Si, Y., Zhang, P., Liu, C., Bartolomei, M., Tang, F., Liu, B., Yu, J. and Lan, Y. (2019) 'Combined Single-Cell Profiling of lncRNAs and Functional Screening Reveals that H19 Is Pivotal for Embryonic Hematopoietic Stem Cell Development', *Cell Stem Cell,* 24(2), pp. 285-298.e5.

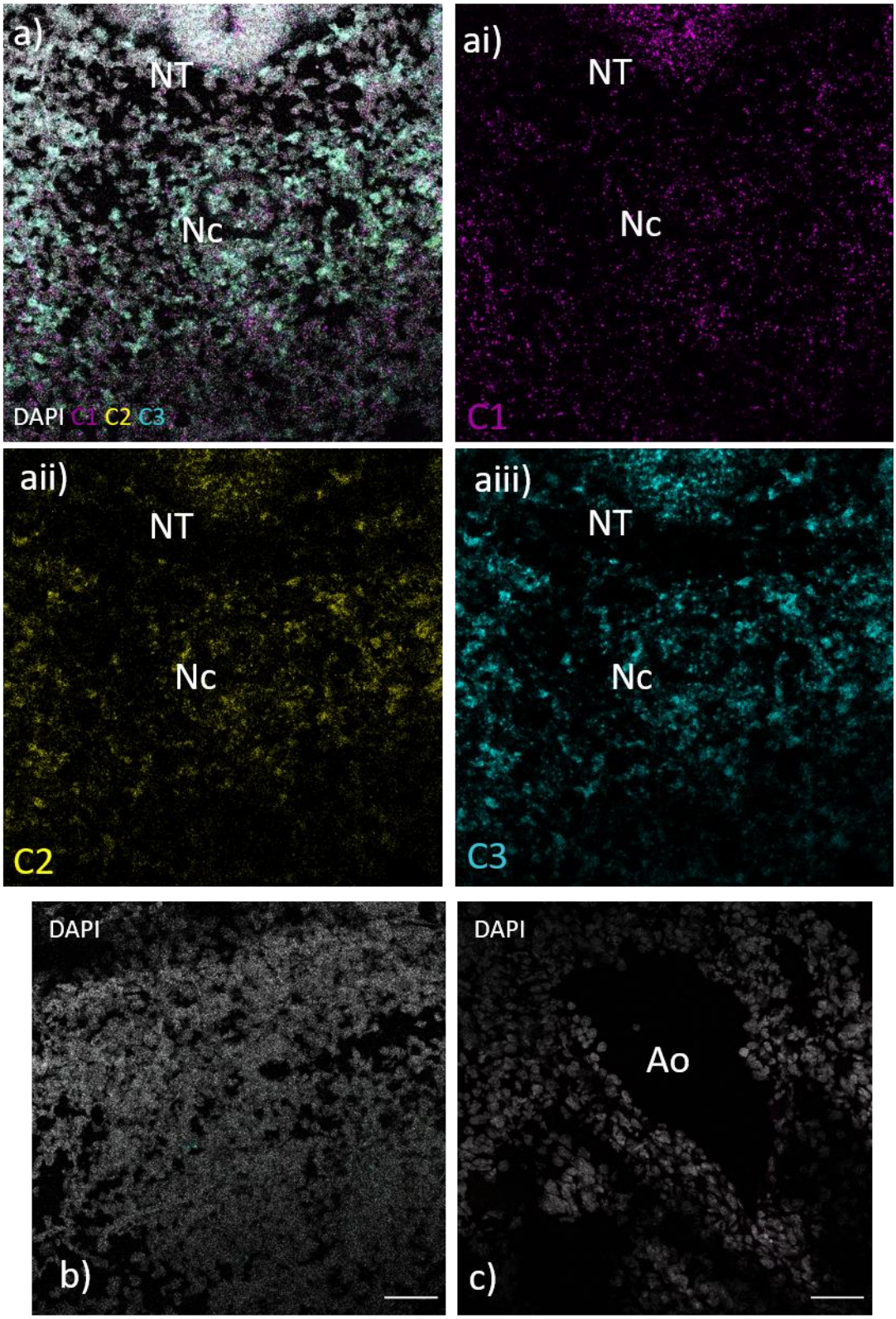
Zohar, R., Suzuki, N., Suzuki, K., Arora, P., Glogauer, M., McCulloch, C. A. and Sodek, J. (2000) 'Intracellular osteopontin is an integral component of the CD44-ERM complex involved in cell migration', *J Cell Physiol,* 184(1), pp. 118-30.

Zovein, A. C., Turlo, K. A., Ponec, R. M., Lynch, M. R., Chen, K. C., Hofmann, J. J., Cox, T. C., Gasson, J. C. and Iruela-Arispe, M. L. (2010) 'Vascular remodeling of the vitelline artery initiates extravascular emergence of hematopoietic clusters', *Blood,* 116(18), pp. 3435-44.

8. Appendix Figures



Appendix 1. CS16 LCM-Seq sample N3_Cut 1 (C1). Immunostaining of VE-Cadherin (VC) and Runx1. This region showed a lack of IAHCs and was not used in the CS16 LCM-Seq differential gene expression analysis. Scale bar = 50µm



Appendix 2 Control slides: ai-iii)) CS16 human transverse section with RNAScope positive controls for each channel (C1-Opal520, C2-Opal570, C3-Opal690). NT = neural tube. Nc = Notochord. b) Negative control for RNAScope. C) Negative control for human CS16 immunostain with secondary antibodies only.